

**NASA CONTRACTOR
REPORT**

NASA CR-577



NASA CR-577

C.1

LOAN COPY: RETURN TO
AFWL (WLIL-2)
KIRTLAND AFB, N MEX

BIOLOGICAL CONTROL SYSTEMS -

A CRITICAL REVIEW AND EVALUATION

*by Lawrence Stark, Laurence R. Young, Robert Taub,
Arthur Taub, and Peter G. Katona*

*Prepared by
BIOSYSTEMS, INC.
Cambridge, Mass.
for Ames Research Center*

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • SEPTEMBER 1966

465



NASA CR-577

BIOLOGICAL CONTROL SYSTEMS -

A CRITICAL REVIEW AND EVALUATION

By Lawrence Stark, Laurence R. Young, Robert Taub,
Arthur Taub, and Peter G. Katona

Distribution of this report is provided in the interest of
information exchange. Responsibility for the contents
resides in the author or organization that prepared it.

Prepared under Contract No. NAS 2-2122 by
BIOSYSTEMS, INC.
Cambridge, Mass.

for Ames Research Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 - Price \$4.25

FOREWORD

This report is part of a comprehensive review and evaluation of research on biological control systems, emphasizing the methods of investigation and state of models proposed for a variety of systems. The contribution of control theory to understanding complex biological feedback systems is stressed and the limitations of both mathematical tools and instrumentation are considered.

This work has been carried out by Biosystems, Inc. since 1963 under contracts NAS 2-1372 and NAS 2-2122 sponsored by NASA Ames Research Center. Mr. Richard Weick was the technical monitor. The review has been supervised by Dr. Laurence R. Young and Dr. Lawrence Stark, with contributions by Drs. A. Taub, R. Taub and P. Katona. A separate portion, subtitled "Developments in Manual Control" by L.R. Young and L. Stark, was published in March 1965 (NASA CR-190). Related material on the visual system was published in NASA CR-238 "Physiology of the Visual Control System", by L. Stark, C. Kupfer and L.R. Young, June 1965.

TABLE OF CONTENTS

SECTION 1	
Introduction - Defining Biological Feedback Control Systems by Lawrence Stark, M.D. and Laurence R. Young, Sc.D.	1
SECTION 2	
Vestibular Control System by Laurence R. Young	15
SECTION 3	
Pupil and Lens Control Systems by Lawrence Stark	97
SECTION 4	
Control Systems and Body Fluid Regulation by Robert Taub, M.D. and Arthur Taub, M.D., Ph.D.	131
SECTION 5	
Control of Sensation from the Skin by Arthur Taub	225
SECTION 6	
The Cardiovascular System by Peter G. Katona, Sc.D.	343

SECTION 1

INTRODUCTION - DEFINING BIOLOGICAL FEEDBACK CONTROL SYSTEMS*

How do biological control systems differ from conventional control systems? At first glance a biological control system looks very much like its electromechanical counterpart, and the engineer is likely to consider it as merely a "flesh and blood variation". Thus the analogy is quickly seen between the pupil servomechanism and the photocell-controlled lens-opening control of an automatic movie camera, since both act to change the size of an opening, and thereby regulate the intensity of light hitting a sensitive surface. The two control systems appear similar when viewed in block diagram form because investigations of biological control systems borrow heavily from the highly developed techniques of feedback control systems analysis.

In the typical research involving a biological servomechanism, the system's input-output characteristics are studied, particularly with regard to its frequency response and transient responses to deterministic inputs. On the basis of this data, physiological information, and a moderate amount of

*This chapter was presented in part to New York Academy of Sciences, June 1963 and published in the Annals of the NYAS Vol. 117, Article 1, pp 426-442, 1964, under the title "Defining Biological Feedback Control Systems".

guessing, a model is constructed to represent the system. This model may be analytical, a computer simulation, or a physical "working model". The assumption is that one may experiment with different inputs, systems parameters, and topographical connections of the model, and thereby learn about the true nature of the normal and pathological functioning of the original control system. This approach was used by Wiener in his construction of a phototropic mechanical "moth" to demonstrate and study parkinsonism and intention tremor (1).

Difficulties encountered by a control engineer when he first attempts to model a control system are many. Chief among them is the inherent existence of a number of characteristics of biological servomechanisms which are not generally present in mechanical counterparts, and therefore likely to be overlooked by the engineer on his first approach.

The general representation of a feedback control system shown in Fig. 1 includes most of the components which the control engineer seeks to identify. The system output, coming from the controlled element (or fixed process), is the variable being controlled to follow an input reference command. The error between the input command and the output is the actuating signal driving the controller, which in turn actuates the controlled element so as to reduce the error. One of the functions of a feedback control system is to reduce the effect

of any unwanted disturbance on the output.

The proper identification of input, output, error, and disturbance forms a necessary first step in analysis of a biological system. There remain, however, a number of other peculiar characteristics common to many biological control systems - characteristics for which the engineer should be alerted. In the block diagram of Fig. 2, some of these features are illustrated.

The input from other senses indicates the possibility of modification of the adaptive controller, as well as the presence of additional inputs. Biological adaptation is the common occurrence of accepting a new steady-state input level as a reference, and changing feedback gain accordingly. The precognitive input predictor reflects the ability of many biological control systems to recognize predictable periodic input signals and use the predictable nature of the signal to compensate for inherent delays in the controller or the controlled element. The task adapter and the adaptive controller represent the ability of many biological control systems to change the nature of control according to the desired characteristics of the task, the type of controlled element, and the makeup of the input signal.

In this introduction the operation of slow biological adaptation, precognitive input prediction, and task adaptation will be demonstrated by three biological control systems -

the pupil reflex to light, the eye movement control system, and the manual tracking servomechanism.

The pupil reflex is a biological control system which has been thoroughly analyzed by use of conventional servomechanism techniques (2). Pupil area changes to regulate the light flux hitting the retina, and the retinal adaptation is included in the feedback path. This system illustrates no predictive behavior, and its frequency response may be determined by straightforward tests with sinusoidal light intensity of different frequencies. Fig. 3 shows the periodic pupil response to such sinusoidal input and indicates a significant phase lag between input and output in an open loop experiment.

Retinal light adaptation plays an important role as an example of biological adaptation in the pupillary servomechanism. For example, when the light level is drastically changed, the pupil contracts or dilates to correct the disturbance. Because of retinal adaptation, however, the pupil redilates or reconstricts slowly to assume almost the same steady area level as before, the very low frequency gain of the pupil reflex to light being quite low. As a consequence, a person may have almost the same pupillary area at two quite different light intensities. If the light intensity is now suddenly changed to some identical intermediate level, the pupil will constrict in the dark-adapted case and dilate in the light-adapted case. When the retina is exposed to darkness for varying periods of time, the sensitivity increases

as a function of time-in-the-dark, and this increased sensitivity is measured as an increased response of the pupil to the same light stimulus (3). It is this process of adjustment of retinal sensitivity to prevailing light flux which acts as a disturbance adaptation in the pupil-reflex-to-light servo-mechanism.

The eye movement control system enables us to direct our eye at many targets by means of rapid saccadic jumps and smooth pursuit movements. The input is a target angle, the output, the angle of gaze, with the error between the two observed at the retina.

Fig. 4 shows what happens when we attempt to investigate the system by means of a simple sinusoidal test signal in the horizontal plane. After an initial delay and several saccadic jumps, the eye locks into synchronism with the periodic input in a clear example of precognitive tracking. To test this predictive apparatus quantitatively, we measure the system gain and phase as a function of frequency for predictable single sinusoids and also for unpredictable sums of noncoherent sinusoids. Fig. 5 shows that the eye moves in close synchronism with the target (0° phase lag and 0 db gain) for predictable sinusoids up to 1 cps. For random signals, however, the eye movement control system shows increasing phase lag and lower gain for increasing frequency, as might be expected of a "follow-up" servomechanism (4).

For its nonpredictive tracking mode there is considerable evidence to indicate that the eye movement control system is a noncontinuous system with independent pursuit and saccadic branches for tracking velocity and position respectively. A sampled data model has been shown to describe the observed frequency and transient responses (5).

As a further test of any model, it is desirable to observe the system response under other feedback conditions than its normal situation of unity feedback. For input-output measurements under normal tracking, the closed-loop nature of the system tends to obscure some of the characteristics of the forward-loop elements. Fortunately, the use of the photoelectric eye movement monitor permits the effective variable feedback to be varied at will.

Discrete oscillations growing into a limit cycle are predicted by the high-gain instability of the model, and are observed experimentally. The eye movement system does not seem to adapt or change its forward-loop characteristics when presented with these unusual conditions of feedback. Thus, while it exhibits precognitive adaptive behavior, it does not show any task adaptation to changes in the controlled element.

Manual tracking refers to the actions of a human operator in manipulating some physical control in response to a visual input. This biological servomechanism has received intense study because of its importance in so many man-machine

interfaces, i.e., driving an auto, aiming an anti-aircraft gun, or controlling the orientation of an airplane.

Early attempts to fit a linear continuous model to the human operator, and more recent efforts employing sampled data models, recognize certain adaptive characteristics of the system, and only quasi-linear ones have been proposed (7,8,9). These models are sets of simple linear systems whose gain and lead or lag terms depend upon the task; specifically, bandwidth of the input spectrum and the nature of the controlled element in the closed-loop operation.

Experiments conducted to investigate the adaptive behavior of the human operator reveal an extremely rapid adaptation under certain circumstances. For a simple compensatory-tracking situation, in which the operator responds by moving a control stick to reduce the displayed error to zero, sudden changes were made in the controlled element following the control stick (10). For the first 0.5 sec. following a polarity reversal, the error diverged because of the positive feedback closed loop. When the subject reverses his own control polarity, however, the error is sharply reduced, and fully adapted tracking under new control conditions is achieved within two seconds of the controlled element reversal.

The manual control system thus exhibits task adaptive control, in terms of changing its control law to achieve desired performance consistent with a variety of controlled

element dynamics.

The three biological control systems discussed above illustrate three adaptive characteristics not generally found in most physical feedback control systems. Although the mathematical analysis techniques of control engineering and communication have proved exceedingly valuable in the investigation of biological servomechanisms, one must proceed cautiously when applying them to biological systems. Biological adaptation, precognitive feedback, and task adaptation are only three of the possible characteristics of biological control systems which must be considered when attempting servoanalysis. The possible influence of inputs from other senses, and the motivation of the subject under certain circumstances, can have a profound effect on the individual control system under study.

Blind application of servomechanism theory to biological problems is likely to lead to confusion. The thoughtful application of servomechanism theory, considering the special characteristics of biological control systems, can lead to important insights into the functioning of the biological system.

In the chapters which follow, the state of understanding of several biological control systems is examined. Of these systems the pupil, the lens and the vestibular system have received a moderate amount of attention from the control point of view. The cardiovascular system has received intensive study, but because of its exceedingly complex multiloop structure

has yielded disappointingly few new insights from these studies. The fluid volume control system and peripheral feedback from the skin are just being opened for research with control techniques.

As companions to this report, attention is called to references 11-13, in which the authors treat the biological control problems of manual control, eye movement control and regulation of intraocular pressure.

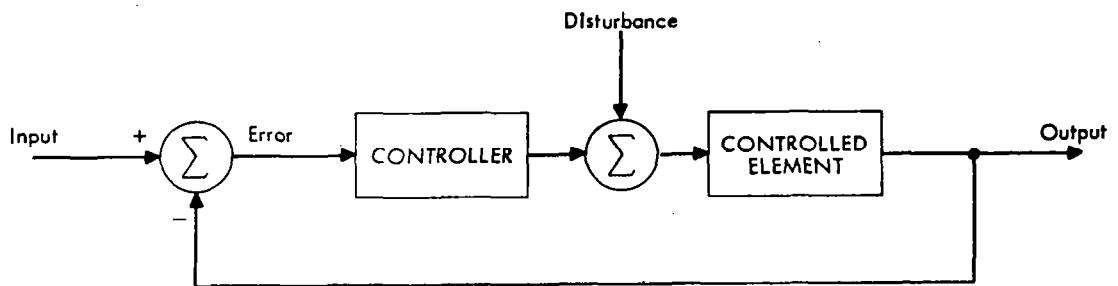


Figure 1. A simple control system.

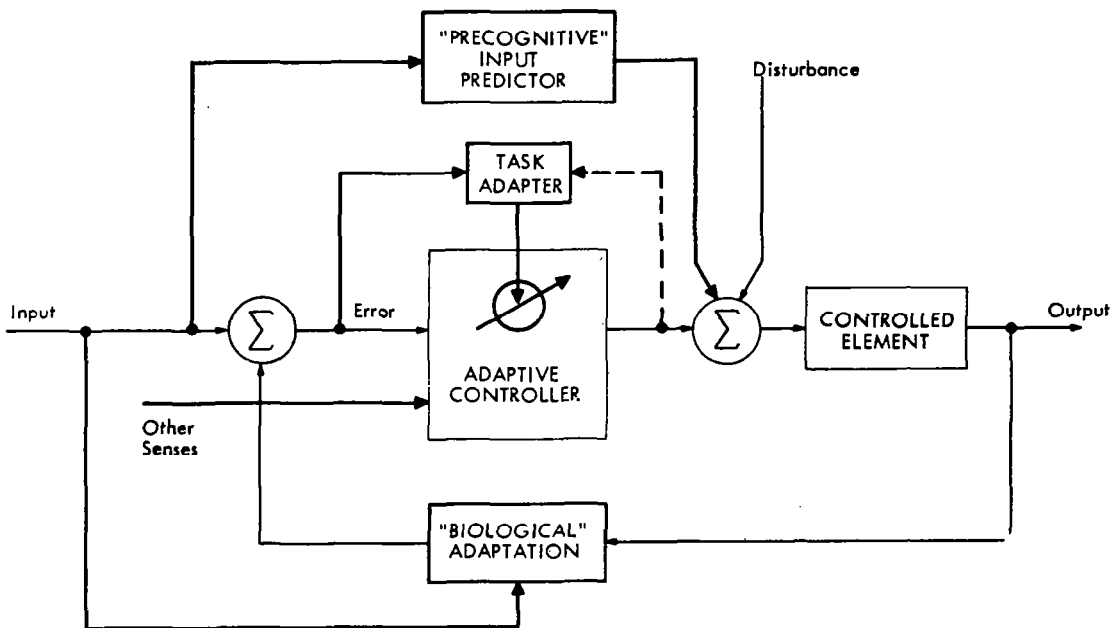


Figure 2. A biological control system.

OPEN LOOP

0.6 Cps

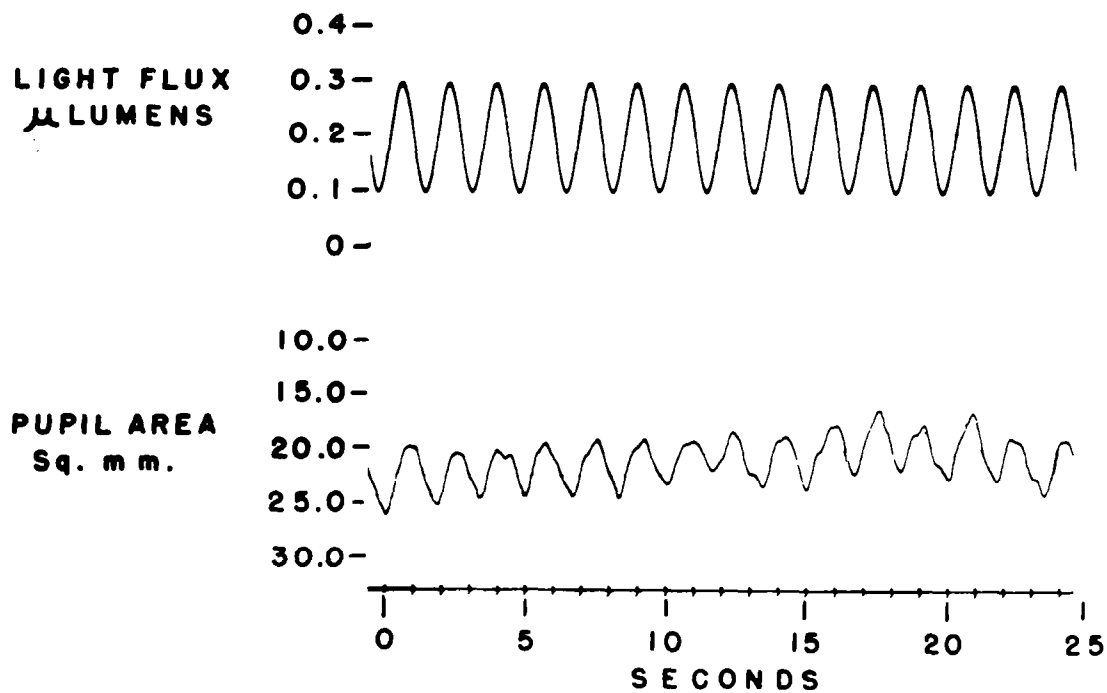


Figure 3. Pupil reflex to sinusoidal light intensity.

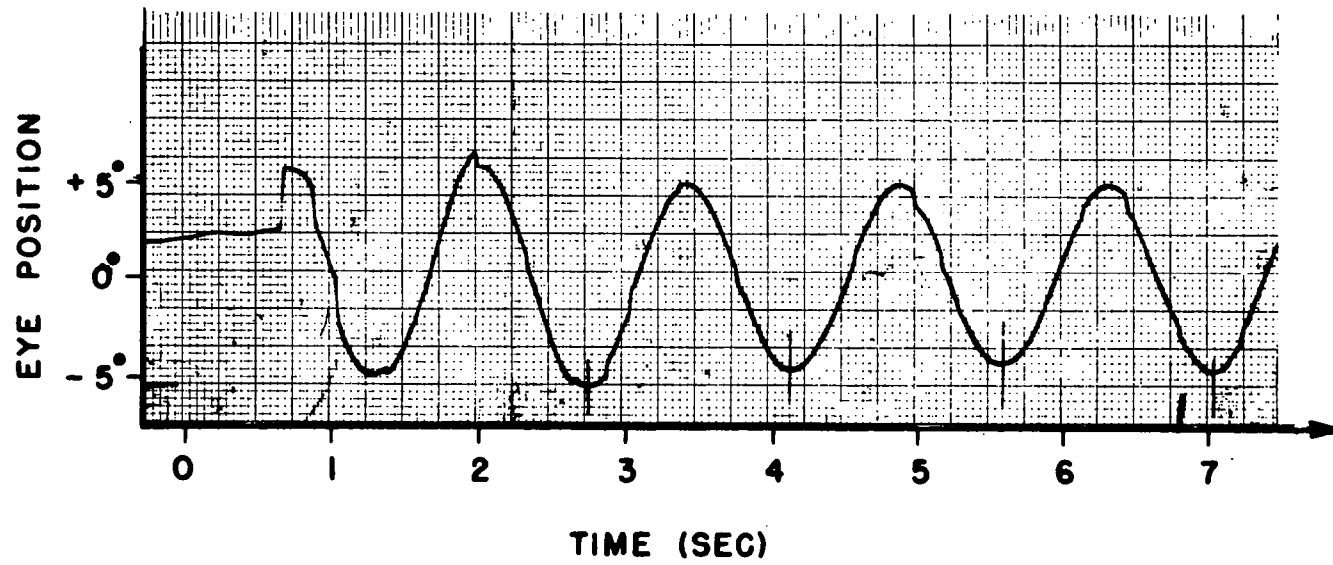
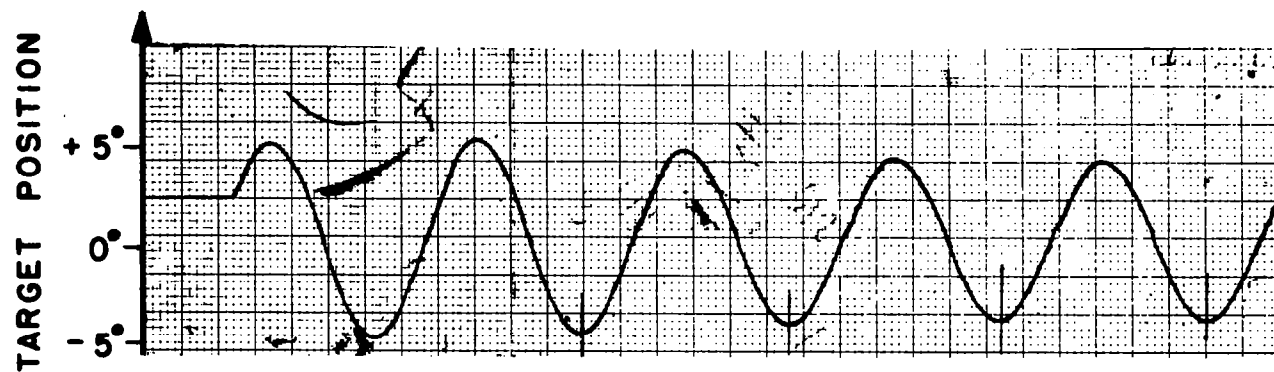


Figure 4. Eye movements following a sinusoid.

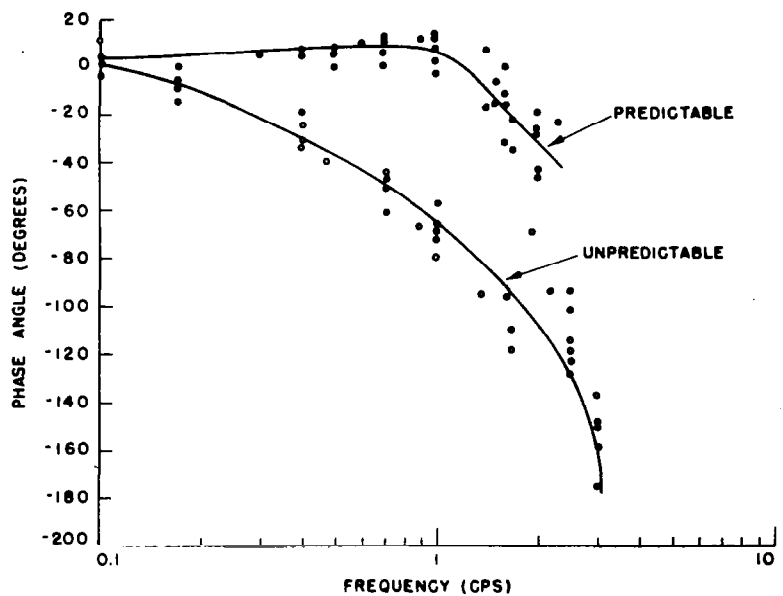
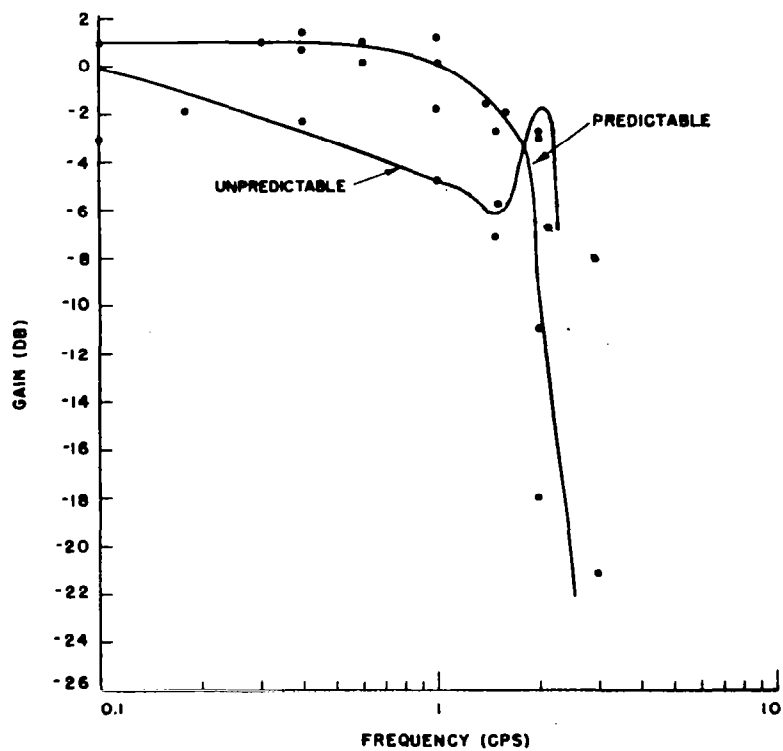


Figure 5. Eye movement frequency response following predictable and unpredictable targets.

BIBLIOGRAPHY

1. Wiener, N. Cybernetics. John Wiley & Sons, New York, N.Y. 1948.
2. Stark, L. Stability, oscillations and noise in the human pupil servomechanism. Proc. IRE 47:1925-1936, 1959.
3. Lowenstein, O. and I. Givner. Pupillary reflex to darkness. Arch. Ophth., 30:603-609, 1943.
4. Stark, L., G. Vossius and L.R. Young. Predictive control of eye tracking movements. IRE Trans. on Human Factors in Electronics, HFE 3(2):52-57, 1962.
5. Young, L.R. A sampled data model for eye tracking movements. Doctoral thesis, Dept. of Aeronautics and Astronautics, M.I.T., Cambridge, Mass. 1962.
6. Young, L.R. and L. Stark. Variable feedback experiments supporting a discrete model for eye tracking movements. IEEE Trans. on Human Factors in Elec. HFE 4(1):38-51, 1963.
7. Tustin, A. The nature of the operator's response in manual control and its implications for controller design. J. IEE 94(11A) No. 2, 1947.
8. McRuer, D.T. and E.S. Krendel. Dynamic response of human operators. WADC Tech. Report, p 56-524, 1957.
9. Elkind, J.I. Characteristics of simple manual control systems. M.I.T. Lincoln Lab. Tech. Rep. III, 1956.
10. Young, L.R., D.M. Green, J.I. Elkind and J.A. Kelly. Adaptive characteristics of manual tracking. 4th Nat'l. Symp. on Human Factors in Elec., Washington, D.C. 1963.
11. Young, L.R. and L. Stark. Biological control systems - A critical review and evaluation. Developments in manual control. NASA CR-190, March, 1965.
12. Stark, L., C. Kupfer and L.R. Young. Physiology of the visual control system. NASA CR-238, June, 1965.
13. Young, L.R. A control systems view of intraocular pressure regulation. Proc. of 6th Int'l. Fed. of Biological Eng. and Med. Electronics, Tokyo, August, 1965.

SECTION 2

VESTIBULAR CONTROL SYSTEM

by
Laurence R. Young

I. DEFINITION OF CONTROL VARIABLES

The vestibular control system in man is one of several systems used to determine his orientation in space. As a single channel of a multi-input control system (other inputs including visual, aural, tactile and kinaesthetic sensations), it is particularly difficult to specify its dynamic characteristics without careful consideration of the other inputs. It is possible, nevertheless, to define a set of inputs and outputs for the vestibular system as if it acted alone.

The vestibular system in man is the nonauditory labyrinth in each inner ear, consisting of sets of semicircular canals and otoliths. The three semicircular canals, oriented in roughly orthogonal planes, respond to angular acceleration about an axis normal to the plane of the canal. The otoliths are stimulated by linear acceleration as well as the gravity field, similar to

three-axis accelerometers. They are actually specific force receivers. (Specific force is the gravity vector minus the linear acceleration vector.)

Since the orientation of the inner ear does not change with respect to the head, we may refer all accelerations to motion of the head. Thus the inputs to the vestibular system are two vector quantities --angular acceleration sensed by the canals and the specific force sensed by the otoliths.

Identification of the output also requires careful consideration. An engineer familiar with inertial navigation systems is tempted to draw the analogy, with the semicircular canals providing the attitude signals usually generated by gyroscopes, and the otoliths providing the linear acceleration signals to be twice integrated, yielding position. Attractive as this speculation may be, there is not evidence that man used these inertial senses for navigation (10). One of the vestibular outputs is a "component" of the perceived orientation in space. Although it is not usual to refer to a perceived quantity as an output of a control system, it is nevertheless true that in the case of the vestibular system, the subjective feeling man receives about his orientation and movement

with respect to an outside reference is an important output. This perceived orientation may be drawn out as an explicit and measureable signal by any of several methods. Simple subjective reports, such as, "I am rotating to the left", or "I am tilted backwards", yield crude information about the direction and time duration of perceived orientation changes. A second class of measurement of changes in perceived orientation uses the well-known techniques of psychophysical quantization of subjective feelings. These tests rely upon the ability of a subject to adjust some measureable physical parameter until it exactly matches his subjective perception of his orientation. The third category of testing the perceived orientation involves tests of the ability of man to control his orientation in a closed loop control system, in which his vestibular system serves as error sensor.

The block diagram representation of the vestibular control system shown in Fig. 1 represents the labyrinthine system from a "black box" point of view as discussed above. The specific force vector \bar{f} , and the angular acceleration vector ($\dot{\omega}_{IA}$) (rate of change of angular velocity of the head with respect to inertial space) are the two inputs. The projection of these vector inputs on the semicircular canals and the otoliths are determined by the orientation

of these elements in the head and the instantaneous orientation of the head with respect to inertial space. The matrix A represents the appropriate linear transformation from an inertially fixed frame of reference to the frame of reference fixed with respect to the subject's head. The output f_b corresponds to the component of specific force along a hypothetical input axis of the otolith, and similarly $(\dot{\omega}_{IA})_b$ represents the component of angular acceleration along the input axis of each semicircular canal. These angular and linear acceleration components are then multiplied by the sensitivity of the otoliths and canals and acted upon by the dynamic response of these organs. Nonlinearities in the organs may be included in the block representation of their sensitivity. The central nervous system is shown to combine the outputs of the otoliths and canals, as well as possible visual, aural, tactile or kinesthetic sensation, to compose a perceived orientation of the man in space. Any adaptation to a new reference orientation in space which cannot be assigned to the sensitivity of dynamic response of the otoliths and canals is arbitrarily assigned to the function of the central nervous system.

An additional output of the vestibular system is the set of control signals sent to the extra-ocular muscles.

It is well-known that the vestibular system, when stimulated, drives the eyes conjugately in the direction opposite to perceived rotation, thereby tending to stabilize the eyes in space despite rotations of the head. The importance of optokinetic nystagmus as an indicator of vestibular function will be discussed in detail later in this report.

When the vestibular sense is used for closed loop control, as in riding a bicycle, flying an airplane, or walking a tightrope, it leads directly to manual control compensation by turning a control wheel or adjusting one's weight for better balance. The resulting motor control output can be measured as an additional indicator of the vestibular function. When the manual response is processed through the control dynamics of the controlled element, it results in a change in the orientation of the controlled element, i.e., the attitude of the airplane or the tilt of the bicycle. For closed loop representation of the functioning of the vestibular control system, the output may be considered as the controlled reference orientation. Feedback to the vestibular system occurs as the orientation of the head in space varies as a direct result of the change in controlled reference orientation.

A great deal of the experimental and theoretical investigations of the vestibular system have been concerned with the normal and abnormal functioning of the portion of the block diagram from input acceleration to the output as perceived orientation or vestibular nystagmus. In its normal operation in man's daily experience the vestibular system serves as an integral part of a closed loop control system through some external controlled element, which may be no more complex than his own postural control system or may involve control of complex vehicles.

In the following sections, we will review and evaluate various studies which bear on the makeup of the boxes shown in this general block diagram. In particular we will deal with the anatomy and physiology of the system, behavioral data relating to primarily psychological observations on its normal functioning, results of linear and nonlinear control system identification attempts to describe the vestibular system, and discussion of the attempted control theory models for its operation. We will conclude with a suggestion of possible further experimental and analytical programs which might uncover some of the hidden aspects of this biological control system.

II. ANATOMY AND PHYSIOLOGY

The vestibular apparatus, comprising the nonauditory portion of the labyrinth in the inner ear, consists of two major portions: the semicircular canals and the vestibule. Figure 2 shows the structural arrangement of the vestibular apparatus and Fig. 3 presents a simplified drawing of the important functional members.

The semi-circular canals, three on each side of the head, lie in planes which are nearly mutually orthogonal, and permit angular accelerations about any axis to be sensed. The horizontal canals lie in a plane tilted up in front by about $25-30^{\circ}$ from the actual horizontal. The entire vestibular apparatus is filled with endolymph fluid with density and viscosity close to that of water.

Each canal starts from a common sac called the utricle, forms roughly a semicircle, and returns to the other end of the utricle. Near the utricle each canal has an enlarged region called the ampulla, in which the sensory elements are located. Figure 4, a cross-section of an ampulla, shows that it is nearly entirely sealed by the cupula and crista. The cupula, which is a gelatinous substance of the same density as the endolymph,

is displaced by movement of the fluid in the canal. The cupula lies above the crista, a rigid sensory cell formation, and is imbedded with cilia-hair-like sensory cell endings from the crista which sense displacements of the cupula. The crista is supplied with approximately 20,000 nerve fibers to transmit vestibular information to the brain.

The cupula hermetically seals the ampulla. When the angular acceleration of the head causes the endolymph fluid to lag behind the structure in a canal, (inertial reaction forces), the cupula is displaced from its normal position. The displaced cupula has its position transmitted through the crista, over which it can be moved, and also exerts an elastic restoring force on the fluid.

Since the cupula is of the same density as the endolymph it is perfectly floated, and therefore does not bend under gravitational force or linear acceleration.

The vestibule, forming the other major portion of the vestibular system, consists of two fluid filled sacs, the utricle and the saccule. Each of these sacs contains an otolith (literally ear stone), a heavy gelatinous mass containing calcium carbonate crystals and resting on the sensory macula (see Fig. 5). The utricle, otolith and

macula are roughly horizontal, and the saccule sensors perpendicular.

Under the action of specific force (gravity or linear acceleration) the heavier otoliths tend to move through the less dense fluid in the direction of the force. The otolith slides across the macula, restrained by supporting strands which limit its movement to about 0.1mm. Just as in the case of the crista and cupula, the macula containing sensory cells has cilia imbedded in the otolith. These cilia are pushed or pulled as the otolith moves in response to specific force, and transmit this information through the macula to the brain by means of the shear force they exert on the sensory cells. The otoliths are sensitive to changes in magnitude as well as direction of specific force.

Coding of information from the vestibular sense organs is of an FM nature. In the normal position of the otolith or cupula the neural message is a resting discharge rate of action potentials. Shifting the otolith in one direction increases the discharge rate, and in the other direction decreases the frequency. Detailed studies of this mechanism have been conducted in the crayfish.

For small deflections around the normal position of the cupula the change in discharge frequency is roughly proportional to the displacement. Naturally the frequency cannot decrease below zero --leading to a saturation in one direction for each canal corresponding to a cupula deflection of the order of 30° .

In addition to the nerve fibers which fire at a rate dependent on position, another type of receptor has been observed in the otolith sensory cells. These receptors are sensitive only to changes in otolith position, and decrease to zero activity when the otolith assumes a new steady-state position. These rate receptors are located near the borders of the macula, with the position receptors occupying the central regions. No such rate receptors have been discovered in the semicircular canals.

The actions of the vestibular mechanisms on the two sides of the head are synergetic --with the activity levels from both sides compared before a single resultant signal is transmitted up the central nervous system. Thus the traumatic loss of one labyrinth causes disorientation which is partially compensated in a period of one year.

III. BEHAVIORAL EXPERIMENTAL RESULTS RELATING TO THE VESTIBULAR CONTROL SYSTEM

The vestibular mechanism plays a varied role in controlling man's actions in his daily activities. The major behavioral implications of its role are reviewed in this section.

Postural Control

The otoliths, and particularly the utricle, are the chief source of non-visual information to the postural control system. They sense the direction of the apparent vertical and control muscular tone of the arms, legs, and neck, etc. to counter the gravitational force and avoid falling over. (When available, visual and tactile cues also provide inputs to the postural control system.) Experiments indicate the inability of cats to right themselves when dropped from an inverted position in the absence of visual cues and after denervation of the utricula macula.

Since the semicircular canals are not affected by linear acceleration, they can serve no role in postural control against the force of gravity. The weight of internal organs apparently does not provide important cues for the regulation of muscle tone.

Compensatory Eye Movements

When the skull is rotated in space the vestibular apparatus acts to stabilize the eyes with respect to the outside environment and thereby stabilize the image of a stationary point on the retina. It does this by moving the eyes slowly in the direction opposite to the movement, apparently without any initial delay. Since the eyes cannot continue to rotate, after reaching a certain deviation from their central position a fast return phase occurs. The combination of this slow sweep and fast return is known as vestibular nystagmus. Use of this nystagmus as an indicator of vestibular sensation is discussed at length in the next section.

Illusions Attributed to the Vestibular System

Since the vestibular system forms one of the important inputs to the human perception of his orientation it may be expected that bizarre stimulation of the vestibular system should yield strange illusions of movement. When these illusions lead to spatial disorientation by a pilot they are of an exceedingly serious nature (32). Nuttall (133) refers to a study which attributes 14% of fatal aircraft accidents to spatial disorientation. Krauss (103) reports "when there is a conflict between the instruments

and normal sensory mechanism, the flight student is very likely to revert to the use of the sensory cues which he has been using all of his life". The most common illusions are reviewed briefly by Fogel (187). He includes descriptions of the visual-g illusion, the autokinetic illusion, the oculogyral illusion, the oculogravic illusion, the non-visual illusion, and the audiogyral illusion. In those in which the vestibular apparatus plays a part, the illusion may be explained in terms of misinterpretation of vestibular sensation, often because of a failure to realize what the orientation or acceleration of the aircraft actually is. When such illusions result in the subjective loss of orientation with respect to the direction of vertical, pilots are said to have vertigo.

Vestibular Stimulation and Motion Sickness

There is no question that motion sickness may be caused by certain patterns of acceleration sensed by the semicircular canals and otoliths. (Destruction of the labyrinth eliminates susceptibility to motion sickness.) There is also general agreement that motion sickness is more likely during a conflict between visual and non-visual senses of orientation (as in an airplane) than when the two agree. Other factors which are supposedly

contributory include odor, temperature, size of the enclosure, repeatability of the motion, disagreeable sights, distasteful food, suggestion of sickness by others, training and adaptability, warning of the motions to expect, and, perhaps most important, individual susceptibility to motion sickness.

For investigators interested in the control characteristics of the vestibular system a study of the conditions leading to motion sickness might appear fruitful as indications of the inability of the system to give suitable information on body motion. Despite a great deal of careful research, incidental observations and anecdotal information, very little has been established about the etiology of motion sickness. There is no agreement, for example, on whether linear or angular acceleration (or both) contribute to motion sickness and vomiting. The state of knowledge about motion sickness is revealed in the proceedings of a symposium held in 1960 to which ten experts contributed . (Symposium on Motion Sickness with Special Reference to Weightlessness 6570th AMRL-TDR-63-25-June 1963.) In his summary statement the symposium moderator states, "there is a feeling that this is a very complex situation in which

there are a number of uncontrolled variables of undetermined importance."

At the present time there seems to be very little to be learned about the control characteristics of the vestibular system from a study of motion sickness.

IV. CONTROL SYSTEM DESCRIPTIONS AND EXPERIMENTS RELATING TO THE VESTIBULAR SYSTEM

This section deals with those experimental results and models which have direct relevance to a control system description of the vestibular mechanism. The results are divided into two groups; those bearing primarily on the semicircular canal system and those pertaining primarily to the otoliths. It will be seen that a wealth of carefully compiled experimental material exists for evaluation of models of the semicircular canals. The lack of careful experimental data on otolith performance reflects the greatly increased difficulty of performing unambiguous experiments on the otoliths.

SEMICIRCULAR CANALS

Historical

The development of current theories of the operation of the semicircular canals dates from the work of Steinhausen (160), who suggested that the operation of each semicircular canal may be viewed as the mechanical action of a torsion pendulum. The moment of inertia of such a pendulum corresponds to the moment of inertia of the fluid ring in the semicircular canal, the damping term results

from the viscous forces as the endolymph flows through the narrow canal, and the elastic restraining force is attributed to the cupula, which is displaced from its neutral position by any movement of the endolymph in the canal. By showing that the cupula seals the ampulla nearly entirely, not permitting free steady-state flow of the endolymph, Steinhausen dismissed the concept of the canal as a Mach canal, and strengthened the plausibility of a torsion pendulum model. The measurement of suitable output variables to test the validity of this model in man posed a great difficulty at first and Steinhausen suggested that the times of occurrence and disappearance of subjective feelings of rotation should be correlated with a minimum threshold deviation of the cupula.

The earliest experimental attempts to ascertain the parameters of the second order equation proposed by Steinhausen were carried out by Van Egmond, Groen and Jongkees, at Utrecht (165). Their schematic diagram of the semicircular canal is shown in Fig. 6.

The differential equation for the angular deviation of the endolymph in relation to the skull, and therefore

the presumed angular deviation of the cupula is given by:

$$\theta \ddot{\xi} + \pi \dot{\xi} + \Delta \xi = \alpha \theta$$

where

θ = moment of inertia of the endolymph,

π = moment of friction at unit angular velocity of the endolymph with respect to the skull,

Δ = stiffness, or torque moment per unit angular deflection of the cupula (ξ),

ξ = angular deviation of the endolymph with respect to the skull,

$\dot{\xi}$ = angular velocity of the endolymph with respect to the skull,

$\ddot{\xi}$ = angular acceleration of the endolymph with respect to the skull,

α = component of angular acceleration of the skull, with respect to inertial space, normal to the plane of the semicircular canal.

The dynamic response of this model is completely determined by two parameters, π/θ and Δ/θ . Most of the experimental results to be reviewed in this section relate to determinations of these constants by indirect methods.

Theoretical Determination of π/θ

An approximate value for the ratio of the friction coefficient to the moment of inertia may be calculated

from anatomical data on the semicircular canals (165). Assuming that the mass of endolymph which must be moved includes an effective continuation of the semicircular canal through the whole circle in the utricle, its moment of inertia is given by

$$\theta = 2\sigma\pi^2r^2R^3$$

where

R = radius of the canal,

σ = density of the endolymph,

r = circular cross section radius of the canal.

In calculating the value of Π , it is assumed that significant frictional moments exist in the narrow portion of the canal but not in the utricle, and therefore only half the circumference of the circle, πr is considered. Application of Poiseuille's Law to this situation yields

$$\Pi = 8\eta\pi^2R^3$$

where

η = the viscosity of the endolymph.

The ratio Π/θ is therefore independent of R, and given by

$$\Pi/\theta = 4 \gamma / \sigma r^2$$

Use of the values $\gamma = 0.006$ (grams/cm/sec), $r = 0.03$ cm, and $\sigma = 1.0$ grams/cm³ yields the approximate value ratio of

$$\Pi/\theta \approx 27 \text{ sec}^{-1}$$

Theoretical determination of the value of Δ/θ , would presuppose a knowledge of the elasticity or spring restraining force of the cupula when deflected by the endolymph. Since such information has never been obtained, no theoretical estimate of this parameter may be offered. Based on indirect evidence, however, the value of Δ/Π appears consistently to yield a level of approximately 0.1 sec^{-1} . Considered with the estimate of Π/θ given above, this would yield a value

$$\Delta/\theta \approx 2.7 \text{ sec}^{-2}$$

Experimental Determination of Torsion Pendulum Parameters

Duration of post-rotation sensation. The roots of the basic torsion pendulum equation are given by

$$\omega_{1,2} = \frac{-\pi \pm (\pi^2 - 4\Delta\theta)}{2\theta}$$

Since the system is very highly damped (the quantity under the square root sign is positive and much greater than one) the roots are widely separated, yielding one very short and one very long time constant. Assuming $\Delta/\pi \ll \pi/\theta$, the roots are

$$\begin{aligned}\omega_1 &\approx -\Delta/\pi \\ \omega_2 &\approx -\pi/\theta\end{aligned}$$

In Laplace transform notation the relationship at cupula position to head acceleration is

$$\frac{\xi(s)}{a(s)} = \left[\frac{1}{(s + \Delta/\pi)(s + \pi/\theta)} \right]$$

Consider what would happen to the cupula if the skull were suddenly brought to a stop after steady-state rotation at constant velocity γ rad/sec. Solution of the equation with initial conditions $\xi = 0$ and $\dot{\xi} = \gamma$, yield

$$\xi = \gamma \frac{\theta}{\pi} (e^{-\frac{\Delta}{\pi} t} - e^{-\frac{\pi}{\theta} t})$$

The resulting deviation is plotted in Fig. 7 where the slow exponential decay appears as a straight line on a plot of log cupula deflection versus time. The physical interpretation of this movement is as follows: Following the sudden stopping of the skull, the angular momentum of the fluid carries it past the zero position with initial velocity γ , and it is slowed to a stop by the friction of the endolymph in the canal; the endolymph and cupula each reaching a maximum deflection. ($\xi_{\max} \approx \gamma \theta / \pi$) time constant for this fast deflection of the cupula is approximately θ / π . Following this maximum deflection the weak elastic force of the cupula slowly forces the endolymph to return to its initial position, opposed chiefly by the friction force. During this period the system resembles a simple first order system, and exhibits exponential decay

$$\xi = \gamma \frac{\theta}{\pi} (1 - e^{-\frac{\Delta}{\pi} t})$$

To make use of this hypothetical relationship for tests of the torsion pendulum parameter, van Egmond et al. assumed the threshold of subjective sensitivity to rotation is equivalent to a specific angular deviation of the cupula ξ_{\min} . Whenever the cupula deflection exceeds this value the subject should report the sensation of rotation. The curve of Fig. 7 shows that the time duration from the stopping of the actual rotation until the cupula falls below its threshold level ξ_{\min} is dependent upon the original angular velocity of the subject.

$$t_u \approx \frac{\pi}{\Delta} \log \left(\frac{\theta \dot{\gamma}}{\pi \xi_{\min}} \right)$$

This equation indicates that the duration of the sensation of rotation should be proportional to the logarithm of the initial angular velocity. Tests to determine this relationship are known as cupulograms, and a typical example is shown by the line marked "sensation" in Fig. 8. The relationship between time duration and the initial angular velocity (called the impulse) is indeed approximately logarithmic. According to the above equation the slope of the cupulogram should yield the value of π/Δ . For the line given in Fig. 8, $\pi/\Delta = 8$ sec, which is about the average found for all normal subjects

tested. The average minimum impulse that leads to any sensation of post-rotation is $\gamma_{\min} = 2.5^{\circ}/\text{sec}$.

Step response to acceleration. The cupulogram discussed above is based on the time necessary for the cupula to return to the threshold level and indicate a cessation of the sensation of rotation. If the subject is started from rest and rotated at a constant acceleration $\alpha \text{ rad/sec}^2$, the angular deviation of the cupula should be

$$\xi \approx \alpha \frac{\theta}{\Delta} (1 - e^{-\Delta t / \pi})$$

This time course is shown in Fig. 9. The subject will first sense rotation when his cupula deviation reaches ξ_{\min} , which occurs after a latency of τ seconds. For small values of τ compared with π/Δ ,

$$\begin{aligned} \xi_{\min} &= \alpha \frac{\theta}{\pi} \tau \quad \text{or} \\ \alpha \tau &= \xi_{\min} \pi / \theta \end{aligned}$$

The prediction would be that the product $\alpha \tau$, which van Egmond et al. call the "Mulder" product should remain constant. Their results show it approximately constant ($1.5\text{--}2.0^{\circ}/\text{sec}$) over angular accelerations from $1\text{--}5^{\circ}/\text{sec}^2$. Moreover, this Mulder product which indicates

the integrated acceleration necessary to force the cupula over to its threshold level, should just be equal to γ_{\min} , the threshold level of initial impulse found from the cupulogram experiment. That value ($\gamma_{\min} = 2.5^{\circ}/\text{sec}$) is in fairly good agreement.

Sinusoidal stimulation. By rotating the subject about a vertical axis on a torsion swing the skull could be forced to undergo sinusoidal accelerations of instantaneous value $a \sin \omega t$. The second order system model for the semicircular canal predicts that ξ should be in phase with the acceleration at very low frequencies, lagging the acceleration input by 90° at the undamped natural frequency ($\omega_0 = \sqrt{\Delta/\theta}$) and finally lagging input acceleration by close to 180° at very high frequencies. By determining the subjective resonance or 90° phase lag point, one can estimate the value of ω_0 . At the resonance frequency the cupula acts as a pure velocity meter, and would indicate to the subject that he is at rest (zero velocity) only at the peaks of his swings on the torsion pendulum. van Egmond et al. find an average value of 1 rad/sec for the natural frequency, thus

$$\Delta/\theta = 1.0 \text{ sec}^{-2}$$

They estimate the probable error in Δ/θ as about 20% and that in π/θ as about 25%.

The torsion swing experiments yield another method of checking the level of ξ_{\min} , the threshold displacement of the cupula. As the torsion swing vibrations gradually decrease in amplitude the test subject will find his sensation decreasing until he senses only the maximum point on the swings (see Fig. 10). Assuming a value of acceleration a_{\min} at a frequency ω to just reach threshold, the cupula equation gives a threshold level of cupula displacement as

$$\xi_{\min} = \frac{a_{\min}}{\omega} \frac{\theta}{\pi}$$

Assuming that the threshold level of the cupula does not change from one type of experiment to another, this yields three ways of checking it. Thus

$$\delta_{\min} = a\tau = \frac{a_{\min}}{\omega} = \xi_{\min} \frac{\pi}{\theta}$$

The three estimates are generally in agreement with an error of less than 25% and fall in the region of 1-2.5°/sec.

Subjective Velocity. In the impulsive stop experiments in which the subject is brought to a sudden halt from a constant velocity rotation, the subjective feeling immediately following the halt is one of continued rotation in the original direction, gradually slowing down and finally decreasing to zero at the threshold time. Assuming that a certain subjective angular velocity corresponds to each position of deviation of the cupula, one could use estimates of subjective angular velocity to indicate the cupula position at any time. As mentioned previously, following an impulsive stop the cupula quickly rises to a level $\xi_{\max} = \gamma \frac{\theta}{\pi}$, and then decreases exponentially toward zero with the time constant π/Δ . Estimates of subjective angular velocity determined from successive estimates of subjective angular position, were established. A typical plot shown in Fig. 11 shows extremely good agreement with predictions. Note first of all that the extrapolated subjective velocity at time zero is almost exactly 40 degrees/sec, which was the true initial angular velocity of the subject. This lends credence to the relationship $\xi_{\max} = \gamma \frac{\theta}{\pi}$ and, since θ/π has been estimated to be 0.1, leads to the conclusion that $\xi = 0.1\gamma$ for human subjects. Secondly, note that the decrease of subjective angular velocity with time follows the expected

exponential decay, and plots as a straight line on the logarithmic plot of the above figure. The slope of this line ($\tau/\Delta = 10 \text{ sec}$) is in good agreement with the estimates obtained from the cupulograms.

Duration of post-rotation nystagmus. The rhythmic motion of the eyes resulting from angular motion of the skull (vestibular nystagmus) may also be taken as an indication of cupula position. In general, the nystagmus consists of a slow phase in which the eyes move slowly in the direction opposite to the subjective rotation, and a quick phase in which the eyes jump rapidly back toward a central position before starting a new slow phase. The angular velocity of the slow phase of nystagmus is thought to be proportional to the deviation of the cupula. (Note that if the cupula position were always proportional to actual skull angular velocity as in the case of a true velocity meter, such an arrangement would permit nearly perfect stabilization of the eyes with respect to the non-rotating environment.)

As an alternate method of determining cupulograms one may measure the duration of the post-rotation nystagmus. Two such cupulograms based on nystagmus are shown in Fig. 8. In general the threshold level of the nystagmus

cupulogram is higher than the sensation level ($5-15^{\circ}/\text{sec}$ as compared to $2^{\circ}/\text{sec}$). This might be interpreted as indicating that the nystagmus threshold corresponds to a somewhat greater deviation of the cupula than does the sensation threshold. It is also to be noted that the nystagmus cupulogram is considerably steeper than the sensation cupulogram, and consequently indicates a higher estimate of Π/Δ . Since the cupula may be expected to return to its steady state position in a unique fashion, presumably following the decay $e^{-\frac{\Delta}{\Pi}t}$, one would expect its position to be reflected by the same rate of decrease of sensation in both subjective and nystagmus cupulograms. The difference in the slope of these two kinds of cupulograms presents a source of possible error in the theory.

Velocity of nystagmus. The angular velocity of the eye during the slow phase of nystagmus may be measured following an impulsive stop to indicate the angular deviation of the cupula. Such a record is shown in Fig. 12. Just as in the case of the subjective estimate of angular velocity mentioned earlier, this plot supports the torsion pendulum theory in two details. The extrapolated initial angular velocity of nystagmus is very close to the true initial angular velocity of the subject.

Furthermore, the decay of angular velocity of the eye is an exponential function of time, leading to an estimate of $\pi/\Delta = 16$ sec. This estimate of π/Δ is consistent with the experimental evidence from the nystagmus cupulogram.

Phase angle of nystagmus velocity. Just as the subjective response to sinusoidal accelerations was used by van Egmond et al. to estimate the system natural frequency, the phase angle of the velocity of the slow phase of nystagmus may be compared to the input angular velocity sinusoid to estimate the phase lag of the cupula at each test frequency. Hixson and Niven (92) present some preliminary results showing the steady state horizontal nystagmus produced by sinusoidal angular accelerations. An approximate curve of nystagmus phase lag with respect to angular acceleration may be deduced from the new data of Fig. 13. By arbitrarily assigning the nystagmus direction reversal at a point midway between fast phases in opposite directions, we calculate the following table:

Frequency (cps)	Phase Lag (degrees)
0.015	28
0.03	44.5
0.075	77.7
0.15	81.5
0.30	95.5

This data is plotted on Fig. 14 against a family of curves for the phase lag of simple second order system with different damping coefficient. The nystagmus phase lag is forced to pass through 90 degrees at the resonance frequency by assuming,

$$\begin{aligned}\omega_o &= 1.5 \text{ rad/sec} \\ (f_o &= 0.24 \text{ cps})\end{aligned}$$

Except for the data point at 0.075 cps the nystagmus phase lag versus frequency resembles that of a second order system with undamped natural frequency of 1.5 rad/sec and damping constant of 3-4.

Assuming a damping constant of 4, the associated second order equation would be

$$\ddot{\xi} + 12 \dot{\xi} + 2.25 = 0$$

This compares to a reasonable degree with the equation parameter proposed by van Egmond et al., of

$$\ddot{\xi} + 10 \dot{\xi} + 1 = 0$$

Detailed experiments of this type at a wide range of frequencies would be very helpful in detailed resolution of the amplitude and phase versus frequency characteristics of the cupula.

Habituation. The experimental data used in construction of the cupulograms described above were generally derived from successive tests on a given subject. The duration of post-rotation sensation following a certain level of impulse would be recorded, and the test then repeated for a higher impulse level until the entire range dictated by the cupulogram was covered. It has been pointed out that repeated determinations of this cupulogram do not all have the same slope, thereby indicating a decreasing value of π/Δ . The apparent time constant of the cupulo-endolymph system would appear to decrease for repeated tests. In discussing this phenomenon, van Egmond et al. (165) were inclined to attribute it to a deformation of the cupula during the high impulse tests at the high end of the cupulogram. Such deformations, they reasoned, would increase the "leak" around the cupula in the ampulla, thereby leading to a system which would have a different time constant.

Cawthorne, Dix, Hallpike and Hood (19) attribute the change in slope of the cupulogram with repeated

testing to a habituation phenomenon, probably occurring centrally. That is, although the cupula goes through the same physical deflection following each impulse, the central nervous system increases the effective threshold upon repeated stimulation. Fig. 15 shows how the successive determination of points on a single cupulogram could yield a curve of much lower slope than that which reflects the true time constant of the cupula, simply by habituation to successive stimuli. They argue that estimates of π/Δ taken from cupulograms must all be lower than the true value. (It will be recalled that the cupulogram parameters were generally 8 to 10 seconds whereas the value expected from theoretical considerations was approximately 27 seconds.)

Oculogyral illusion. In an effort to overcome the habituation problem in determining π/Δ , Cawthorne et al. made use of the oculogyral illusion as an indicator of cupula position. The oculogyral illusion, as described by Graybiel and Hupp (69), is an illusion in which a visual reference point, rotating with the subject, will appear to move relative to the subject in the direction of rotation. It was attributed by Graybiel to movement of the image of the target across the retina during the slow phase of nystagmus. Thus toward the beginning of a slow rotation to the left, the semicircular canals indicate

a velocity to the left and cause the eyes to move in nystagmus with the compensatory slow phase to the right, resulting in a relative movement of the target to the left with respect to the eyes during the slow phase. At present it is not clear whether the oculogyral illusion is, indeed, completely explained by the slow phase of nystagmus. Nevertheless, the velocity of the illusory movement may be taken as an indication of subjective velocity, and therefore an indication of cupula position. Cawthorne et al. used a psychophysical method to track the time course of oculogyral illusion following sudden cessation of rotation. The subject caused the target to assume a velocity which made it appear stationary to him. The magnitude of this cancelling velocity at all times could be taken as equal to the velocity of the oculogyral illusion.

A typical time course of the oculogyral illusion following a sudden halt is shown in Fig. 16. Note that it does show the rapid rise and slow exponential decay associated with the hypothetical cupula movement following such a stimulus. The rapid oscillations probably reflect the tracking technique of the psychophysical method, rather than any oscillatory cupula movement. Using curves of this type the decay constant of the cupula (π/Δ)

could be determined on a single test, without introducing the habituation difficulties of the normal cupulogram. A series of four such tests are shown in Fig. 17. The slope of these curves indicates a parameter value of

$$\pi/\Delta = 24.$$

This value is more than twice that obtained from sensation cupulograms, and quite close to the theoretical estimate. The effect of habituation is clearly seen as the magnitude of the oculogyral illusion is shown to decrease with successive stimuli.

Caloric stimulation. An entirely different means of stimulating semicircular canals involves the use of caloric stimulation. By pouring warm water into one ear, the temperature of the endolymph at a point closest to the water is raised, and convection currents are set up in the canal. These currents presumably act as a torque on the cupula and cause it to deflect just as would be the case following rotation. The direction of vestibular nystagmus observed following caloric stimulation agrees with such an interpretation. Tests using warm and cold water in the two ears show that the action of the semicircular canals is synergetic, and that

the resultant vestibular rotation signal, as reflected by the direction of nystagmus, is the difference in cupula deflection of the two semicircular canals. The caloric tests are quite convenient and useful for clinical investigations of possible vestibular damage, but are not easily interpreted in terms of control theory descriptions of the system.

Threshold of semicircular canals to rotation. It has been assumed that the threshold of sensation or nystagmus corresponds to a minimum deviation of the cupula (ξ_{\min}). As discussed earlier, van Egmond et al., using impulse stimulations of sudden stopping of the subject from an initial angular velocity, found an average threshold level of $\gamma_{\min} = 2.5$ degrees. Their torsion pendulum model indicates that this threshold impulse should be related to the threshold deviation of the cupula by the ratio of damping constant to moment of inertia.

$$\gamma_{\min} = \frac{\pi}{\theta} \xi_{\min}$$

They show further that the Mulder product ($\alpha \tau$) necessary to reach sensation of rotation during constant acceleration, is approximately equal to γ_{\min} as expected.

A slightly different test involves the minimum angular acceleration which can be sensed regardless of duration of that acceleration. The torsion pendulum model for a step of acceleration yields

$$a_{\min} = \frac{\Delta}{\theta} \zeta_{\min}$$

In the case of constant acceleration notice that the spring constant of the cupula enters the equation, but the friction coefficient does not. Assuming that ζ_{\min} is the same for both types of tests, we should find

$$\gamma_{\min} = \frac{\pi}{\theta} \frac{\theta}{\Delta} a_{\min}$$

Using the values θ/Δ approximately equal to 1, as given by van Egmond, the newer values $\frac{\pi}{\theta} = \frac{\pi}{\Delta} = 24$ found by Cawthorne et al., and $\gamma_{\min} = 2.5^\circ/\text{sec}$, one would expect to find a threshold level of angular acceleration of a_{\min} approximately equal to 0.1 deg/sec^2 .

In the most sensitive tests reported to date, using the oculogyral illusion as an indicator, Graybiel, Kerr and Bartley (71) report a threshold for acceleration changes in both positive and negative directions of 0.12 deg/sec^2 . This is in very good agreement with the

expected value from the torsion pendulum model and the impulse threshold data.

OTOLITHS

General

The function of the otoliths in the vestibular system has long been recognized as providing sensitivity to gravity and linear acceleration. The mass of the otolith, which is considerably more dense than that of the surrounding fluid, is acted upon by inertial forces and displaced in the direction of the specific force acting on the body, thereby shifting its position over the macula and indicating the direction of the specific force. This function has repeatedly been shown on birds and sea animals. It is not clear, however, whether the magnitude of the net specific force is also signaled by the otoliths. The otoliths may be able to indicate the direction of the vertical, for example, but not to measure the force of gravity. If this were the case they would not be true vector sensors, but could be considered to be transducers which yield a unit vector directed along the true specific force vector.

By the very nature of the otolith-macula-endolymph configuration, its dynamic behavior should resemble that of a second order system. The inertial mass corresponds to that of the otolith, the spring constant to the supporting hairs from the macula, and the damping to the viscous force between the otolith and endolymph. One would expect considerably higher natural frequency and lower damping constant than for the larger semicircular canals. Almost no fundamental research has been undertaken, however, to describe the function of the otolith or experimentally determine the values of the parameters which affect its operation. Probably the major reason for this situation stems from the difficulty of performing experiments in which the otoliths are stimulated and the semicircular canals are not.

Static Behavior - Perception of the Vertical

One of the principal tasks of the otolith portion of the vestibular system is the indication of the direction of vertical for postural control. In a long series of experiments Asch and Witkin (5,6,177,178) investigated the factors which bear on human perception of the upright. Of particular interest to this study was their series of experiments on the ability of

subjects to perceive the direction of vertical without any visual cues, while they were tilted to various orientations (177). They found that judgments of the vertical were very accurate when the head and body are upright, with mean errors of the order of 2 degrees. When the head or body was tilted from the erect position, however, the ability to judge the vertical on the basis of non-visual cues was found to be severely hampered. The errors increase with the amount of body tilt, reaching a maximum when the body is in a horizontal position. (Other investigations show that errors are still greater for inverted subjects.) The mean range for judgments of the vertical when the subject was tilted 28 degrees from the true vertical was approximately 10 degrees.

Adaptation Effects in Perception of the Vertical

Do we "forget" the true vertical when we have been exposed to a new orientation in space? Adaptation effects in perception of the vertical in the absence of visual cues were investigated by Passey and Guedry (136). Their subjects were placed in a Link trainer and at various initial attitudes, and asked to return themselves to "straight and level". They found that the final orientations were consistently skewed toward the direction of the initial attitude. The amount of adaptation is of

the order of 60% of the initial angular displacement. Passey and Guedry state, "Following exposure to tilt for a period of sixty seconds, readjustments to the gravitational vertical are significantly less accurate than under immediate readjustment, and the number of errors in the direction of initial inclination is significantly greater under a condition of exposure to inclination, giving evidence of adaptation."

It was mentioned, in the discussion of the anatomy and physiology of the otolith, that nerve endings were discovered which respond only to changes in otolith position on the macula. They might function in a manner to permit the central nervous system a degree of adaptation to any steady state direction of specific force. If this were the case control theory description of the otolith would include a term of the form

$$\frac{1}{TS + 1}$$

where T represents the characteristic time of the adaptation.

Tracking the Direction of the Apparent Vertical in the Absence of Visual Cues

The direction of the apparent vertical, or vector sum of gravity and linear acceleration, may be changed by rotating the subject in a 1 "g" field, or by keeping the subject upright and subjecting him to linear acceleration. Clark and Graybiel (30) used an approximation to the latter method to determine the subjective perception of vertical in the absence of visual cues when the direction of the apparent vertical is varied. They placed a subject on a rotating centrifuge. The slowly varying radial acceleration from the centrifuge, added to the normal 1 "g" field, produced a change in the direction of apparent vertical up to 36 degrees from the true vertical. The subjects attempted to maintain a line at the subjective horizontal during the experiments. A typical tracking result is shown in Fig. 18. The data shows the subjective estimate of orientation to be quite accurate for angles of less than 10 degrees between body and the apparent vertical, whereas for angles of greater than 30 degrees the subjects tended to overestimate the angle of tilt. When the same tests were repeated with a long period of constant radial acceleration, no significant differences were found in the ability of the

subject to estimate the horizontal after the rotation had been decreased to zero. These results are in agreement with those of Witkin and Asch, showing greater errors in perception of the vertical for larger angles of tilt, but do not bear out the findings of Passey and Guedry showing significant adaptation when exposed to a non-vertical orientation. The explanation for this difference probably lies in the time taken to vary the tilt of the subject. Clark and Graybiel conjecture that the adaptation effect in perception of orientation only comes into play if the orientation is changed rapidly but not if the direction of the resultant force with respect to the body is changed slowly.

"Step Response" Stimulation of the Otoliths

Using the same test procedure as mentioned above, Graybiel and Brown (65) investigated the delay in reorientation of the perceived vertical when the direction of the apparent vertical was changing rapidly. This was accomplished as before, placing the subject on a centrifuge and quickly accelerating it up to constant velocity, thereby suddenly changing the direction of specific force. The result of the subject's estimation of the horizontal without visual cues are shown in Fig. 19. This experiment

might be interpreted as a possible step response to stimulation of the otoliths, in which the direction of specific force is changed quickly. Interpreted in this light, we would find an extraordinarily long time constant (of the order of 25 seconds) following the increase in α . A puzzling result is the lack of symmetry in returning toward $\alpha = 0$ as the centrifuge stops and the apparent horizontal comes into agreement with the true horizontal. Rather than attribute these asymmetric step responses to some nonlinearity in the otolith, and attribute characteristic time of 25 to 30 secs to the otoliths, we prefer to interpret this experiment as indicating the effect of subject "set" and cross coupling from stimulation of the semicircular canals. Since the subjects knew the experimental conditions and were aware of their rotation in the centrifuge by stimulation of the semicircular canals, they could not be expected to immediately assume that the new direction of apparent vertical was the "true vertical". The long time constant must therefore be interpreted as a combination of vestibular and tactile input information as well as the complex "reorientation" process going on in the central nervous system.

Flotation Experiments

In all of the experiments described above, the subjects received tactile information through the pressure on their skin. Experiments in well-padded seats show greater variability in estimation of the direction of vertical than when nonambiguous tactile cues were present, thereby indicating that the tactile cues were by no means insignificant. Tactile cues can be eliminated by flotation of the subject in a pool of water, where the pressure over the entire body is approximately constant. Since the internal organs are not floated they will continue to respond to the specific force, but it may be assumed that the primary internal specific force sensor is the vestibular system. Under conditions of flotation in the absence of visual cues, subjects are as likely to swim down toward the bottom of the pool as not, when directed to swim toward the surface. The inability to correctly judge the vertical when in any orientation except those quite close to the erect position is clearly shown by such experiments.

Sinusoidal Stimulation and Threshold of Perception

Using a simple swing apparatus to produce sinusoidal linear accelerations, Walsh (170) investigated the

threshold level of subjective sensation of linear motion. By using several stimulus frequencies he could determine whether such thresholds were dependent upon linear displacement, velocity, acceleration or jerk. As expected, threshold is acceleration dependent and is of the approximate value 10 cm/sec^2 or about 0.01 "g" . These thresholds were all determined for the subject lying down on a stretcher, and do not indicate what the threshold level would be in the erect position. Of further interest in this test are the phase relationships between subjective perception of movement and the actual movement as a function of frequency. Subjects reported the direction in which they thought they were moving. At 1 cps oscillations, the subjective estimate was generally correct and in phase with the actual movement. At $1/3$ and $1/9$ cps, however, the subjective estimate was approximately 90 degrees ahead of the actual movement and the subject would indicate that he was moving in a certain direction when the swing had just come to the end point prior to moving in that direction.

If the assumption is made that the subject interprets the output of the otolith as indicating the acceleration of his motion, this bit of data may be used in estimating the frequency response of the system. His subjective

response is in phase with the maximum acceleration at 1/3 and 1/9 cps. At 1 cps, however, he is in phase with the maximum velocity and thereby indicates a 90 degree phase lag of the system. If only first order characteristics are considered, the break frequency of the system would fall between 1/3 and 1 cps. Assuming a break frequency of 0.5 cps ($\omega_0 = 3.1$ rad/sec) a time constant of approximately 0.3 seconds could be assigned to the otolith system. Such a time constant would be much more in keeping with the expected behavior of the system from a mechanical point of view.

V. CONTROL MODELS FOR THE VESTIBULAR SYSTEM

On the basis of the experimental data discussed in the previous section control models for the functioning of the semicircular canals and otoliths may be constructed. The block diagrams of Fig. 20 reflect those characteristics which are known to be present in the vestibular mechanism. Considering first the block diagram representation of the semicircular canals, the input is assumed to be angular acceleration of the skull with respect to inertial space ($\ddot{\omega}_{IA}$). The matrix transformation $[A]$ projects the inertial angular acceleration vector along the input axes of the three semicircular canals. The semicircular canal dynamics relate the output (ξ) of the cupula to the input angular acceleration in terms of a highly damped second order model. Approximate values for the break frequencies are

$$\Delta/\pi = 0.04 - 0.1 \text{ rad/sec}$$

$$\pi/\theta = 10 - 25 \text{ rad/sec}$$

The threshold level of cupula displacement for sensation is about

$$\xi_{\min} = 0.1 - 0.2 \text{ degrees}$$

under conditions of no adaptation to rotation.

The conversion from cupula displacement to pulse frequency introduces a saturation for large negative accelerations or negative cupula displacements for each semicircular canal; however, the value of S_{neg} has not been determined. The block marked "central habituation" represents the long time adaptation to successive stimulation of the semicircular canals, as shown by cupulogram tests. T_H is a long time constant, probably on the order of hours or days.

The threshold adaptation block is assumed to increase the minimum threshold for sensitivity to sensation of rotation and occurrence of nystagmus following a history of angular acceleration. The output of the system is assumed to be subjective sensation of angular velocity.

Over most of the spectrum of head movements encountered in normal activity the semicircular canal system does indeed act to give indications which represent velocity rather than acceleration. Using slightly different values of the parameters in the second order model for the semicircular canals, Mayne (120) calculated the frequency response shown in Fig. 21. (Note that over

the region 10 to 400 cycles/minute the output of the system in terms of cupula displacement is indeed proportional to the input velocity.) The difference of opinion often voiced on this matter stems from a misunderstanding of the functioning of a second order system. A mass-spring-dashpot system is not an accelerometer or a velocity meter or a vibrometer per se, but only acts as a sensor of one of these elements in a certain frequency range. For a heavily damped second order system such as the semicircular canals, the large middle range over which its phase lag is approximately 90 degrees makes it particularly useful as a velocity meter for these frequencies.

The principal remaining control tasks in the investigation of the semicircular canal involve the magnitude of the threshold, the time course of central and threshold habituation, and the effectiveness of the semicircular canal system incorporated as an active error sensor in a closed loop system.

The otolith block diagram shown in Fig. 20b reflects what little is known about the otoliths to date. The specific force input of gravity minus acceleration is resolved by the orientation of the otolith with respect

to the skull and the skull with respect to inertial space to yield a set of specific force outputs f_b acting on the otoliths. The dynamic characteristics of the otolith-macula system in the utricle are represented by the mass M , spring constant K , and damping C of the mechanical arrangement. No data is available on the magnitude of these parameters, however. The presence of cells whose output is proportional to the change in position of the otolith rather than in its position is represented by the box "position & rate sensitivity". Once again, no detailed investigation of T_L , the lead time constant, has been conducted.

It is known that the otoliths are primarily sensitive to changes in orientation away from the erect position, and are not particularly sensitive to changes in orientation when the subject is tilted at a large angle. This phenomenon is represented by the saturation curve in the block labeled "directional sensitivity". The otoliths also have a threshold to linear acceleration of about 0.01 "g".

Quite obviously the interpretation of otolith outputs in the central nervous system is effected by the presence of other outputs from the semicircular canals, and

the subjective linear velocity and orientation with respect to the vertical must be considered as resulting from both systems. Experiments have yet to be carried out in which the otolith control characteristics are carefully investigated in the absence of semicircular canal stimulation.

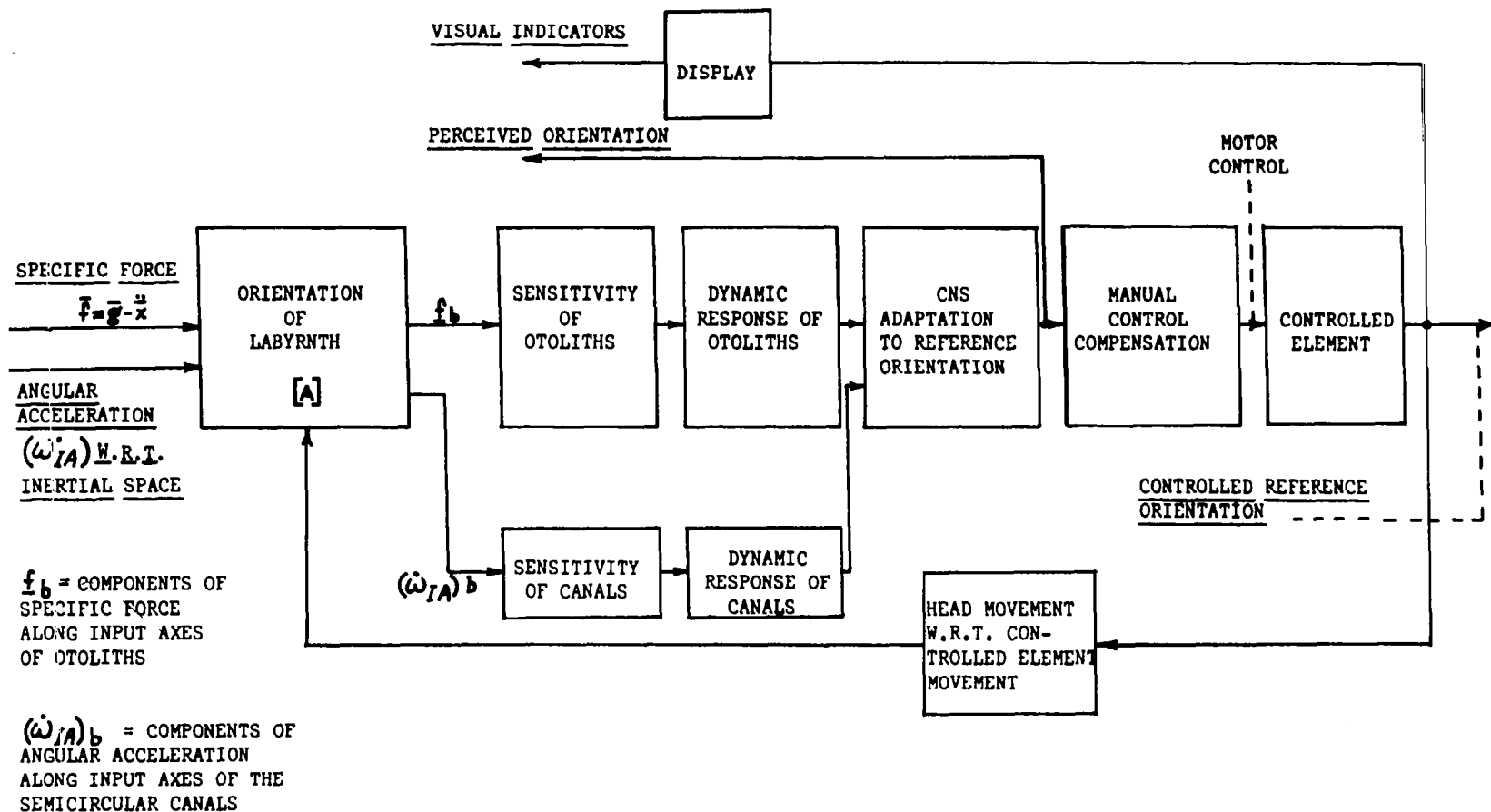


FIG. 1. BLOCK REPRESENTATION OF THE VESTIBULAR CONTROL SYSTEM

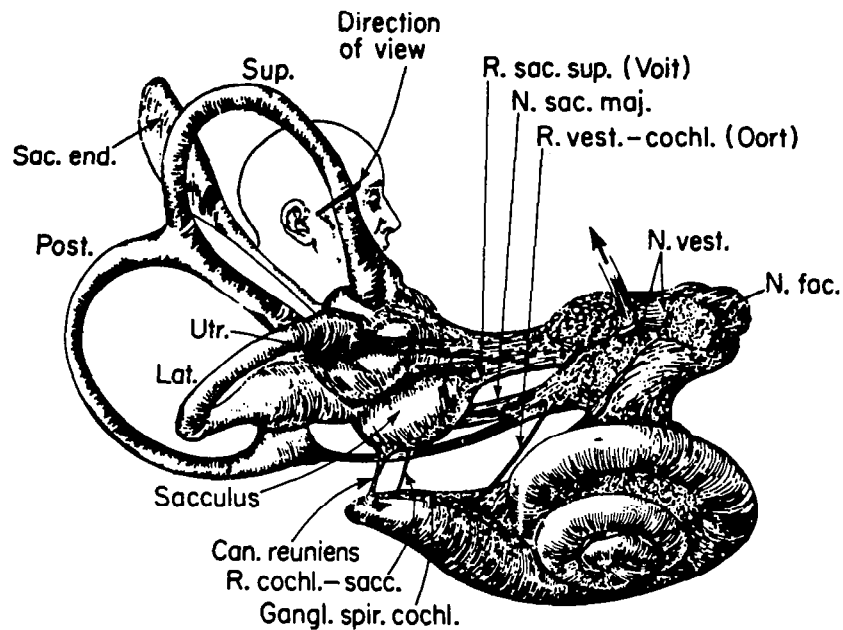


Figure 2. The vestibular apparatus, cochlea, and structural relations of innervation of human labyrinth. (187)

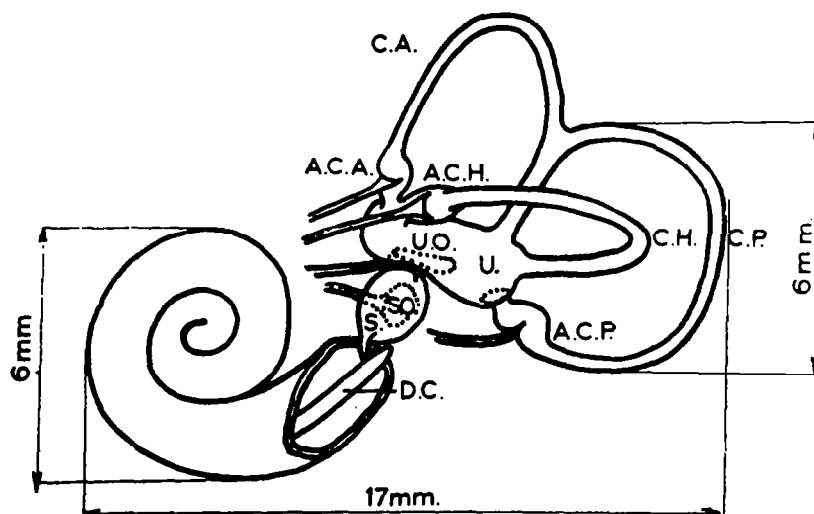


Figure 3. Ventro-lateral view of the left human labyrinth. A., A.C.H., A.C.P., ampullae of anterior, horizontal and posterior canal. D.C., cochlear duct. S. and S.O., saccule and its otolith. U. and U.O., utricle and its otolith. (73)

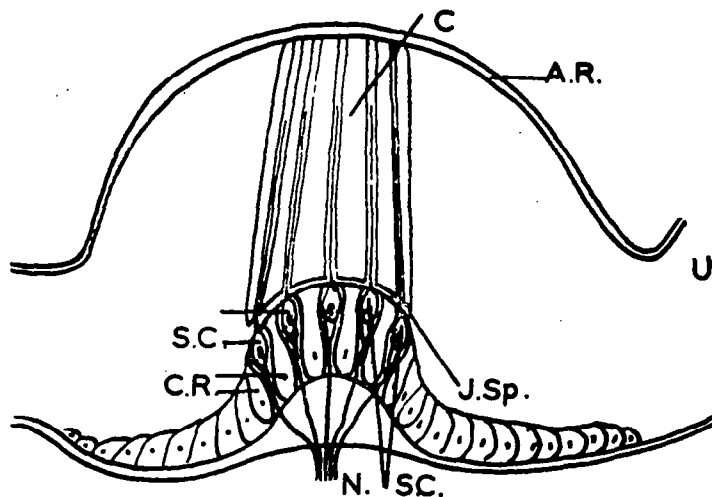


Figure 4. Schematic drawing of a cross section through an ampulla, with cupula and crista. A.R., ampulla roof. C.R., crista, consisting of S.C. (sensory cells), supporting cells and N. (nerve fibres). Between cupula and crista there is the I.S.P. (intercupular space). U., utricle. (73)

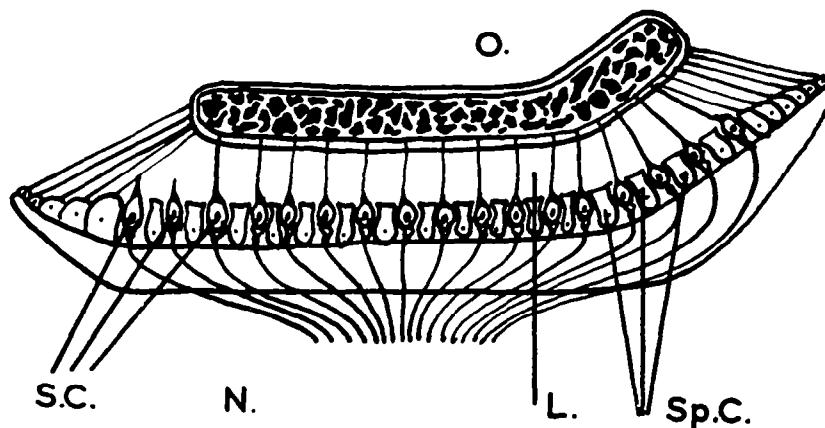


Figure 5. Schematic drawing of a cross section of an otolith and its macula. O. is the otolith, suspended by strands which run from the margins to the macula, consisting of supporting cells (Sp.c.) and sensory cells S.C. Between the otolith and the macula there is a thin layer (L.) to allow the otolith to slide over the macula. N. is the nerve. (73)

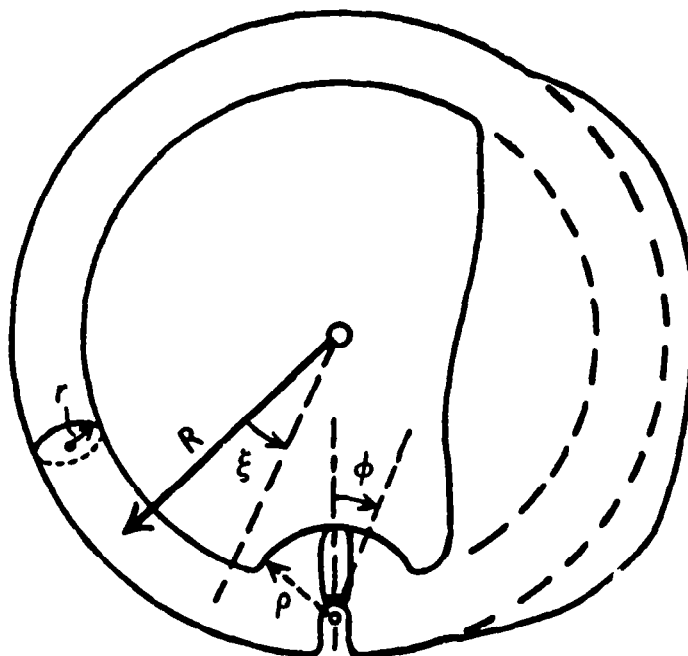


Figure 6. Schematic diagram of the semicircular canal. When the endolymph moves over an angle of ξ the cupula is forced over an angle of ϕ ; ξ and ϕ are of the same order of magnitude.
(165)

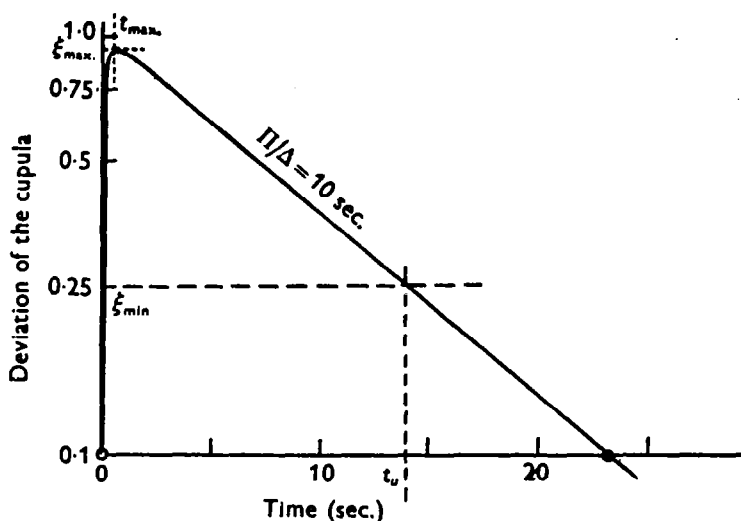


Figure 7. Deviation ξ of the cupula produced by sudden arrest of an angular velocity γ°/sec . Maximum deviation at $t=0.5\text{sec}$.; minimum deviation passed at $t=14\text{sec}$., giving rise to sensation. Ordinate logarithmic. Slope of curve gives Π/Δ . (165)

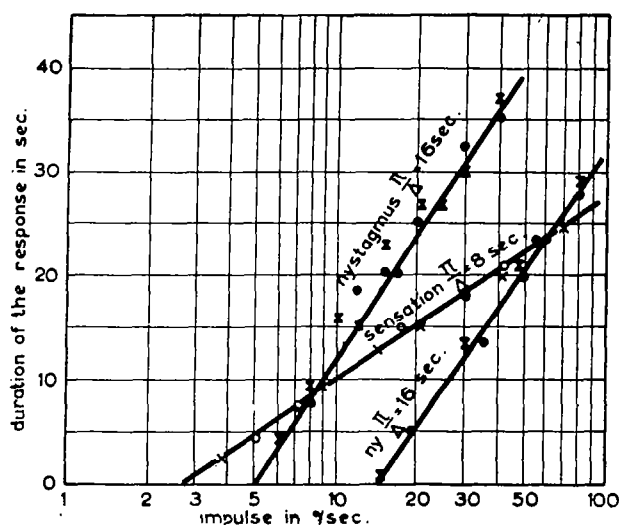


Figure 8. Normal cupulograms of sensation and nystagmus. Depending on the observer's skill, the nystagmus threshold will be localized between 5 and $15^\circ/\text{sec}$; the steepness of the nystagmus cupulograms should stay the same.

O clockwise rotation } sensation ● } idem for (73)
 X counter clockwise } X } nystagmus

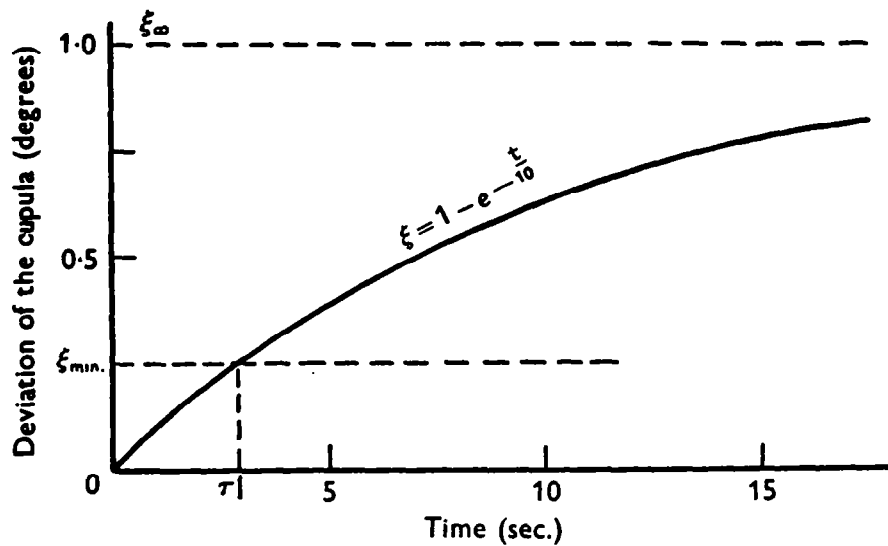


Figure 9. Deviation ξ of the cupula under influence of a constant angular acceleration $(1^\circ/\text{sec.})^2$. The latent period τ seconds lies between start and the moment when the cupula passes ξ_{\min} . After infinite time (about 30 sec. in practice) the cupula attains equilibrium. (165)

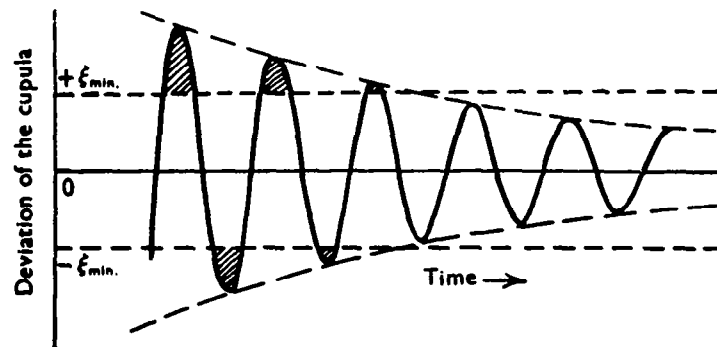


Figure 10. Deviation of the cupula on the torsion swing in the neighborhood of the minimum. Only the tops of the declining sine wave are associated with sensation. (165)

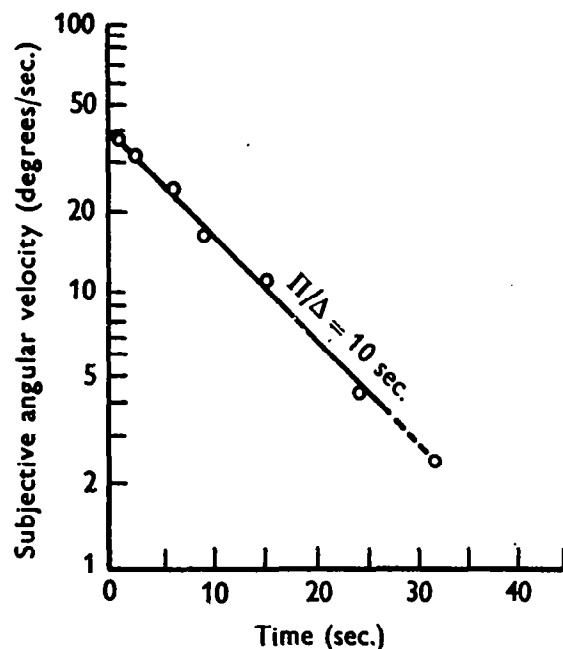


Figure 11. Subjective angular velocity against time. The impulse administered was $40^\circ/\text{sec.}$, as is the extrapolated subjective velocity. The slope gives $\Pi/\Delta=10 \text{ sec.}$ (165)

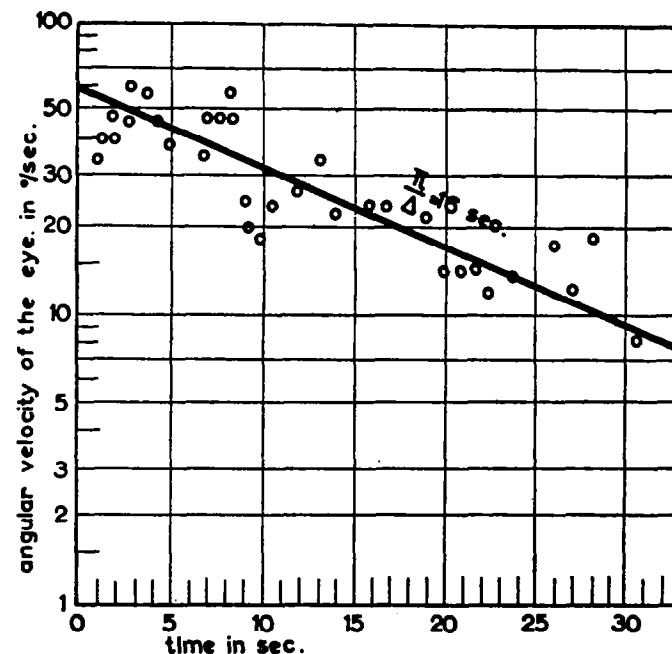


Figure 12. Angular velocity of the eye during the slow phase of the nystagmus (in darkness) as a function of time following an impulse of $60^\circ/\text{sec.}$ The points are calculated from the recording. The extrapolation for $t=0$ gives a value of $59^\circ/\text{sec}$ which appears to be almost equal to the original value. (73)

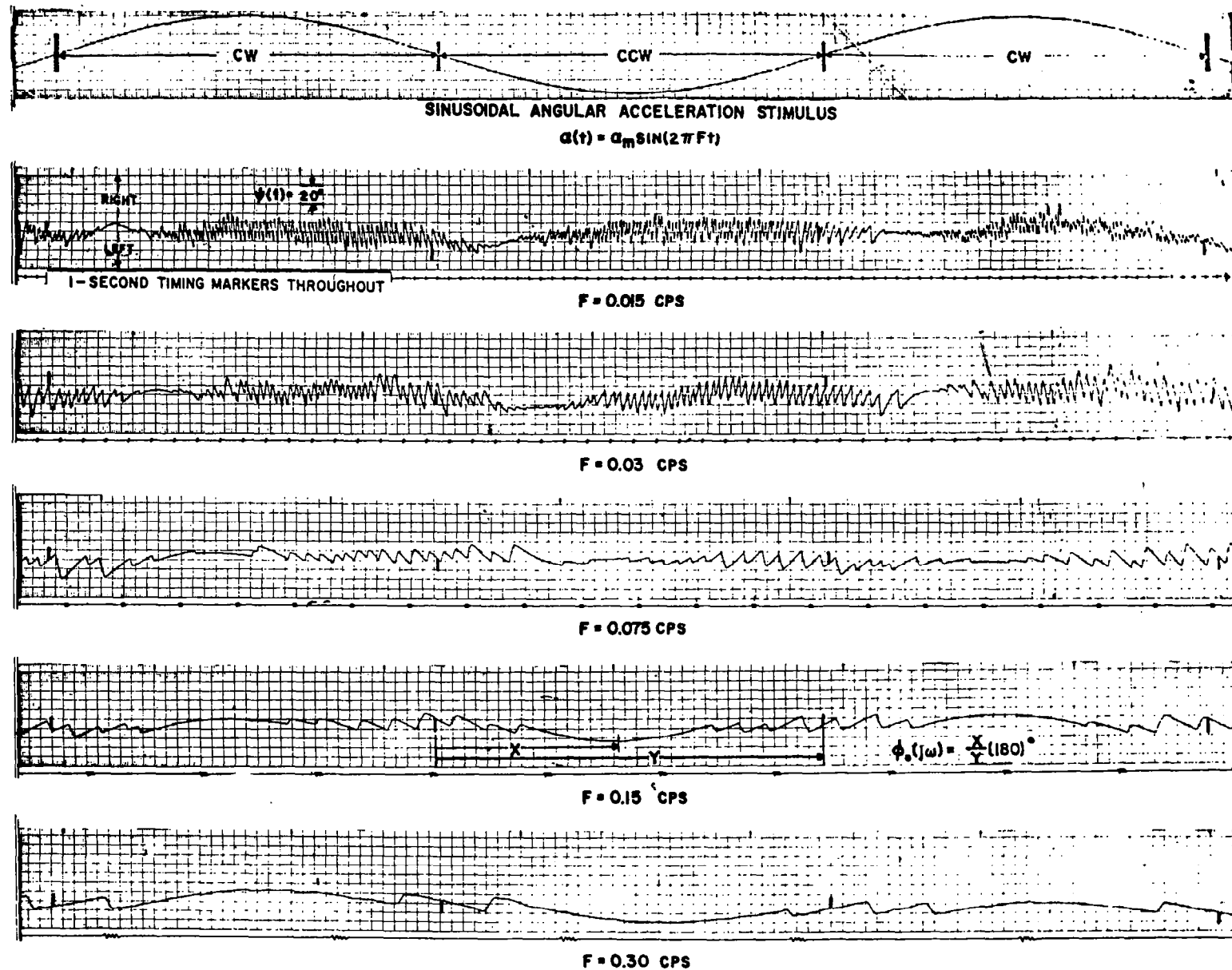


Figure 13. STEADY STATE HORIZONTAL NYSTAGMUS RESPONSES AS PRODUCED BY SINUSOIDAL ANGULAR ACCELERATIONS OF VARIABLE FREQUENCY APPLIED ABOUT THE VERTICAL AXIS

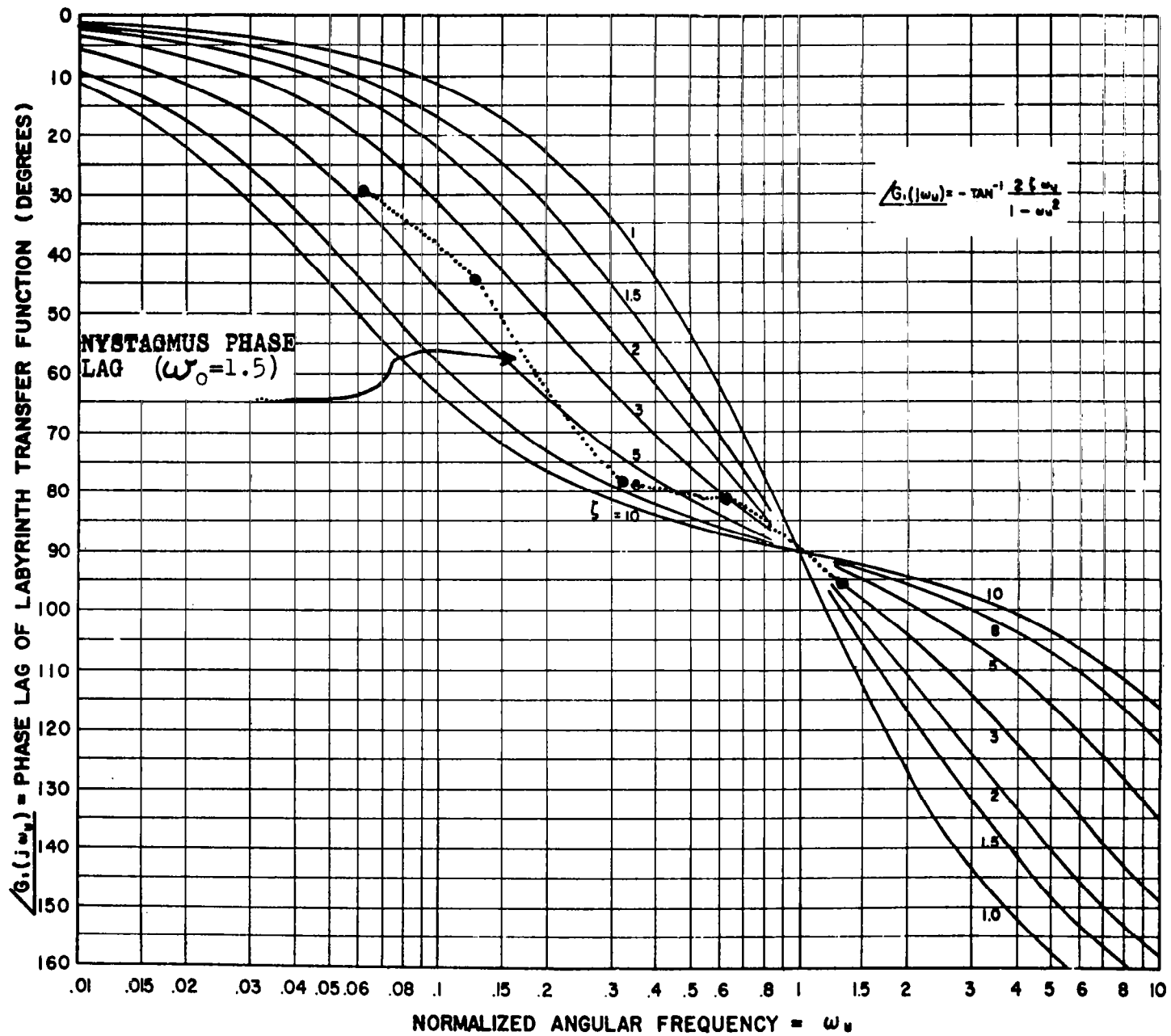


Figure 14. (92)

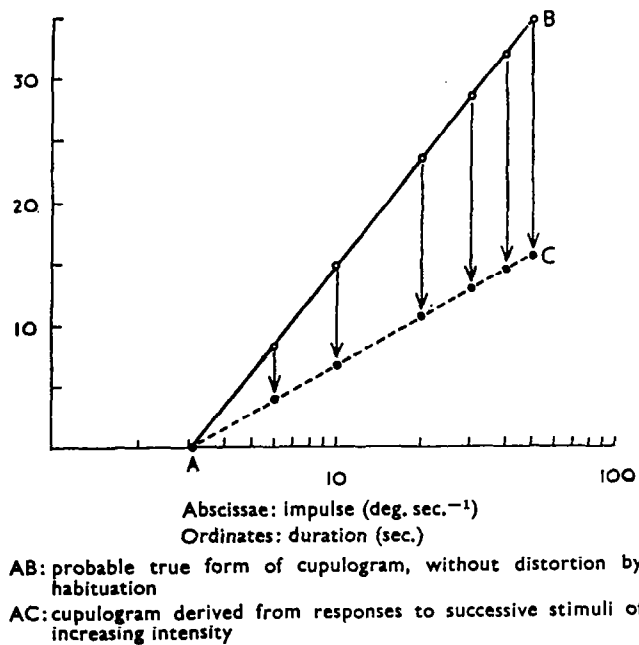


Figure 15. Distorting effect of habituation upon slope of cupulogram. (19)

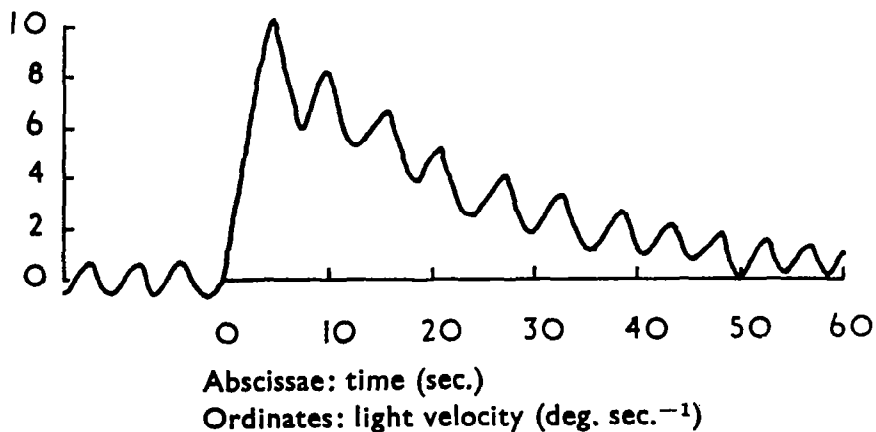


Figure 16. Typical record of subjective response to an impulsive stimulus of $40^{\circ}\text{sec}^{-1}$. (19)

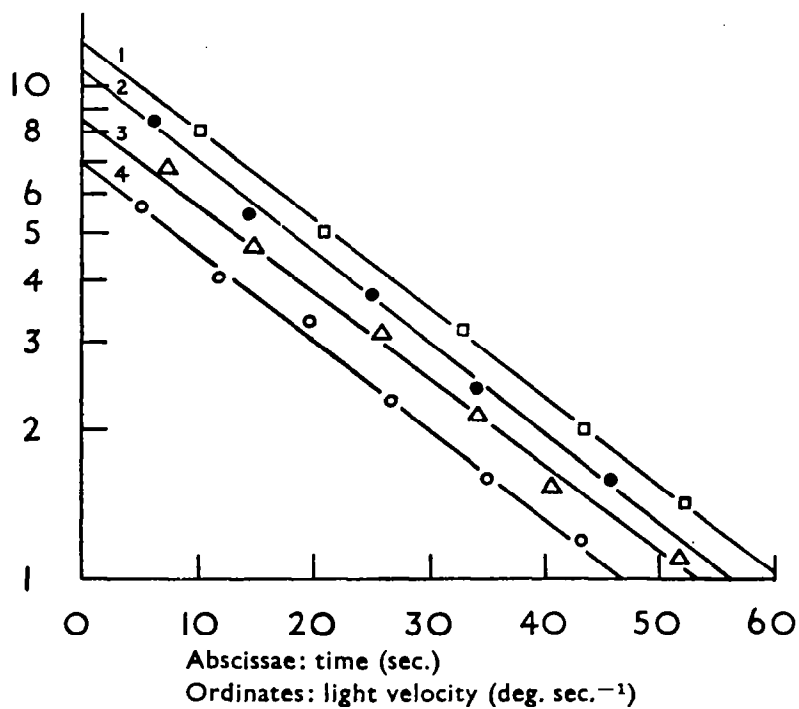


Figure 17. Responses to four impulsive stimuli of $40^{\circ}\text{sec}^{-1}$ applied in succession. (19)

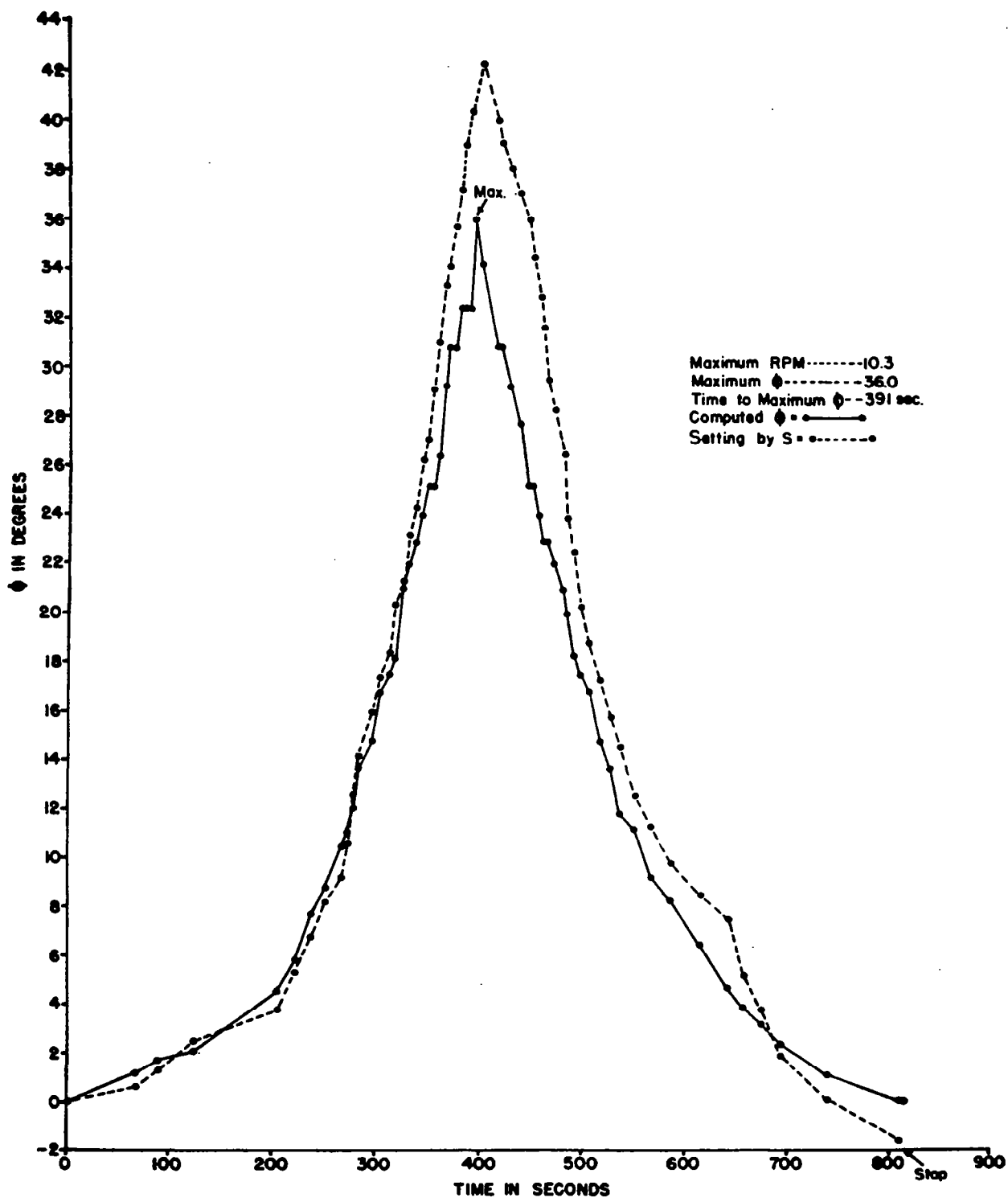


Figure 18. The computed value of ϕ and the setting of the line by subject S during a single trial in Part I. (30)

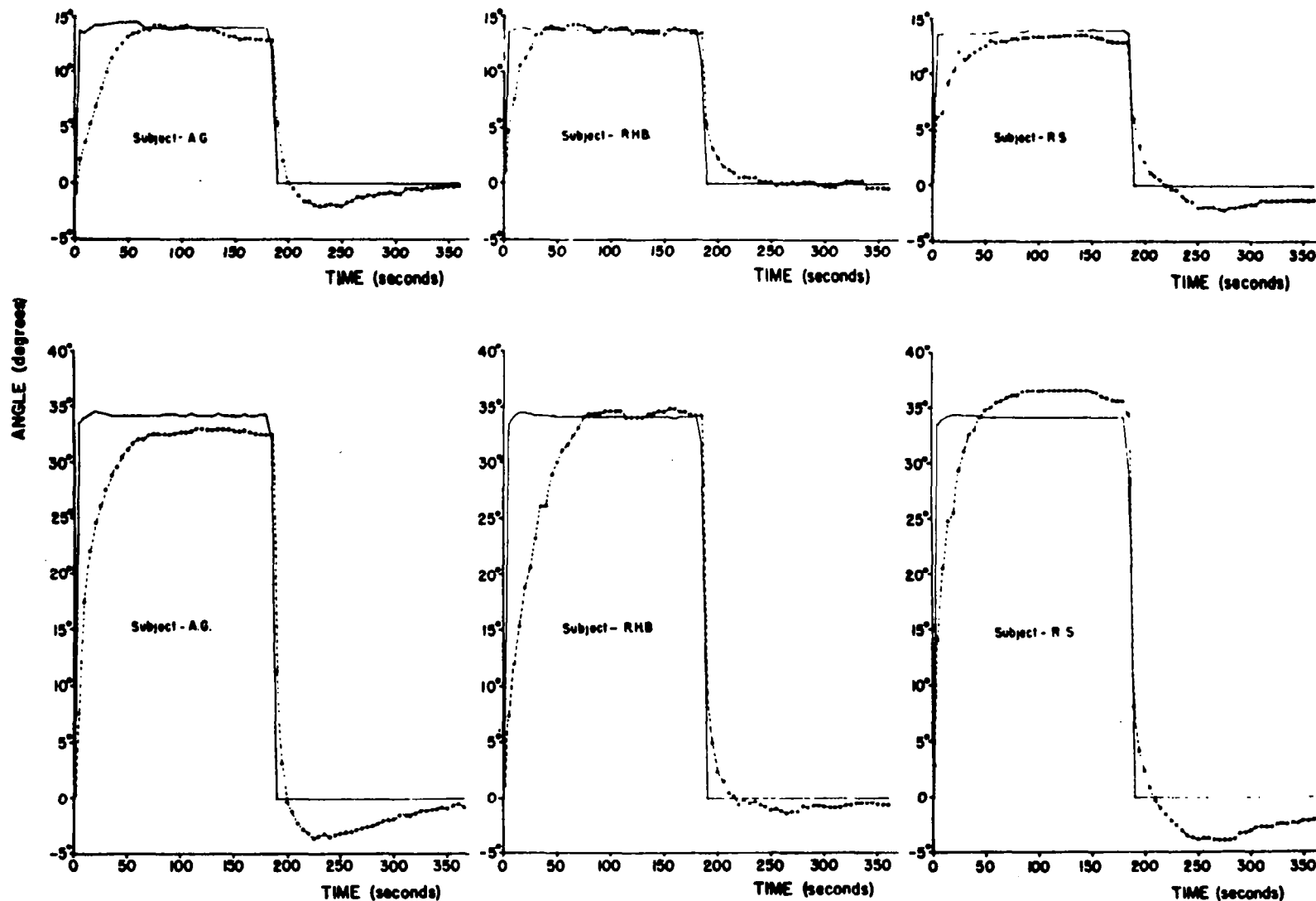


FIGURE 19

SHIFT IN THE DIRECTION OF RESULTANT FORCE AND OF VISUAL ORIENTATION FOLLOWING ONSET OF CENTRIFUGAL FORCE

The solid curve indicates the deviation of resultant force from gravitational vertical. The broken curve represents the change in the subject's estimate of the horizontal. Each point is the average of 10 measurements on separate trials for each of the three subjects. (ref. 65)

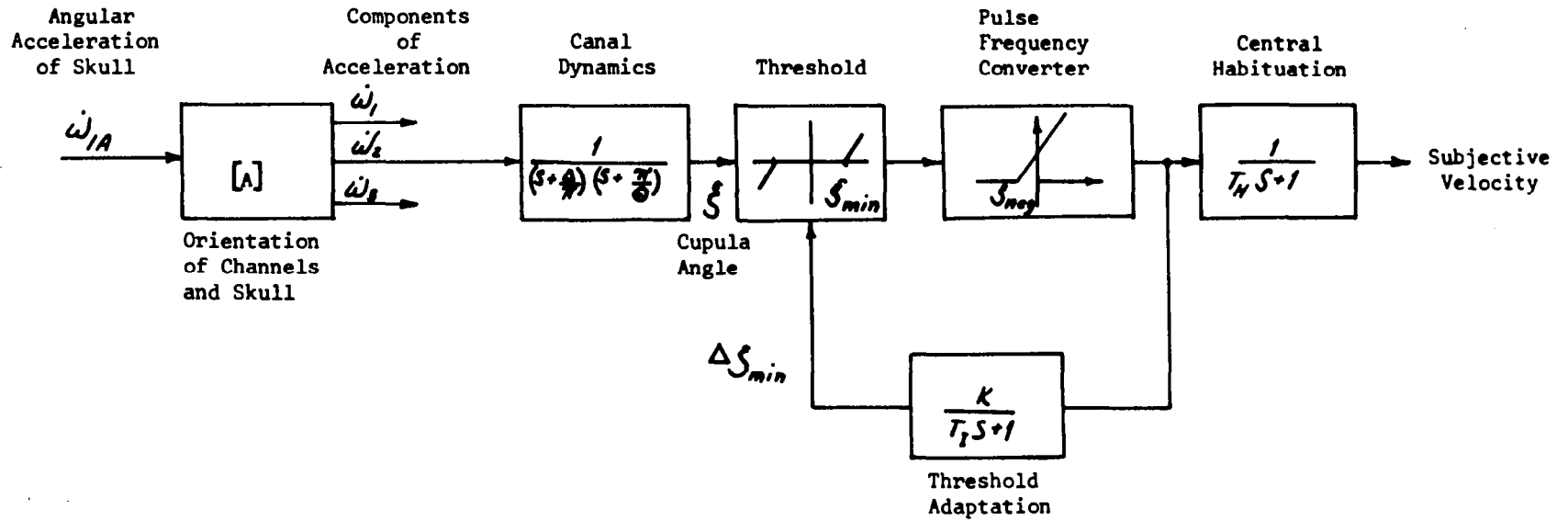


Fig. 20(a) Semicircular Canal Block Diagram

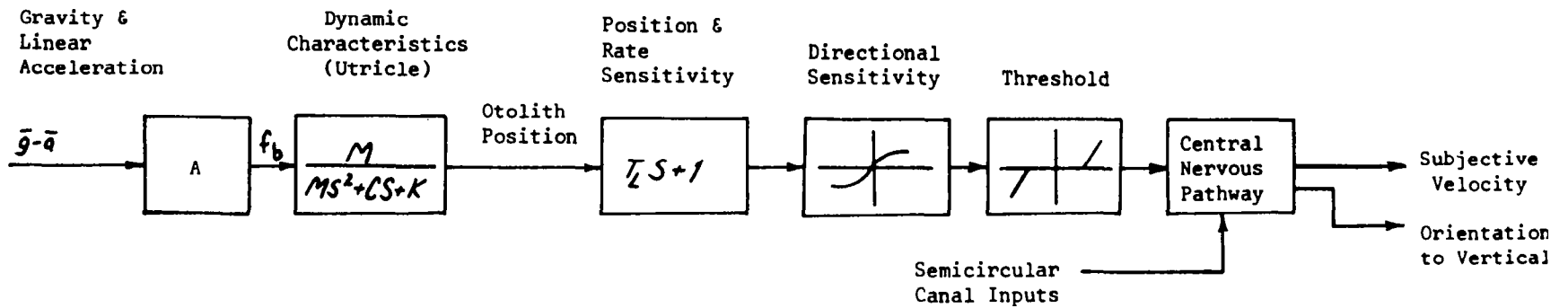


Fig. 20(b) Otoliths Block Diagram

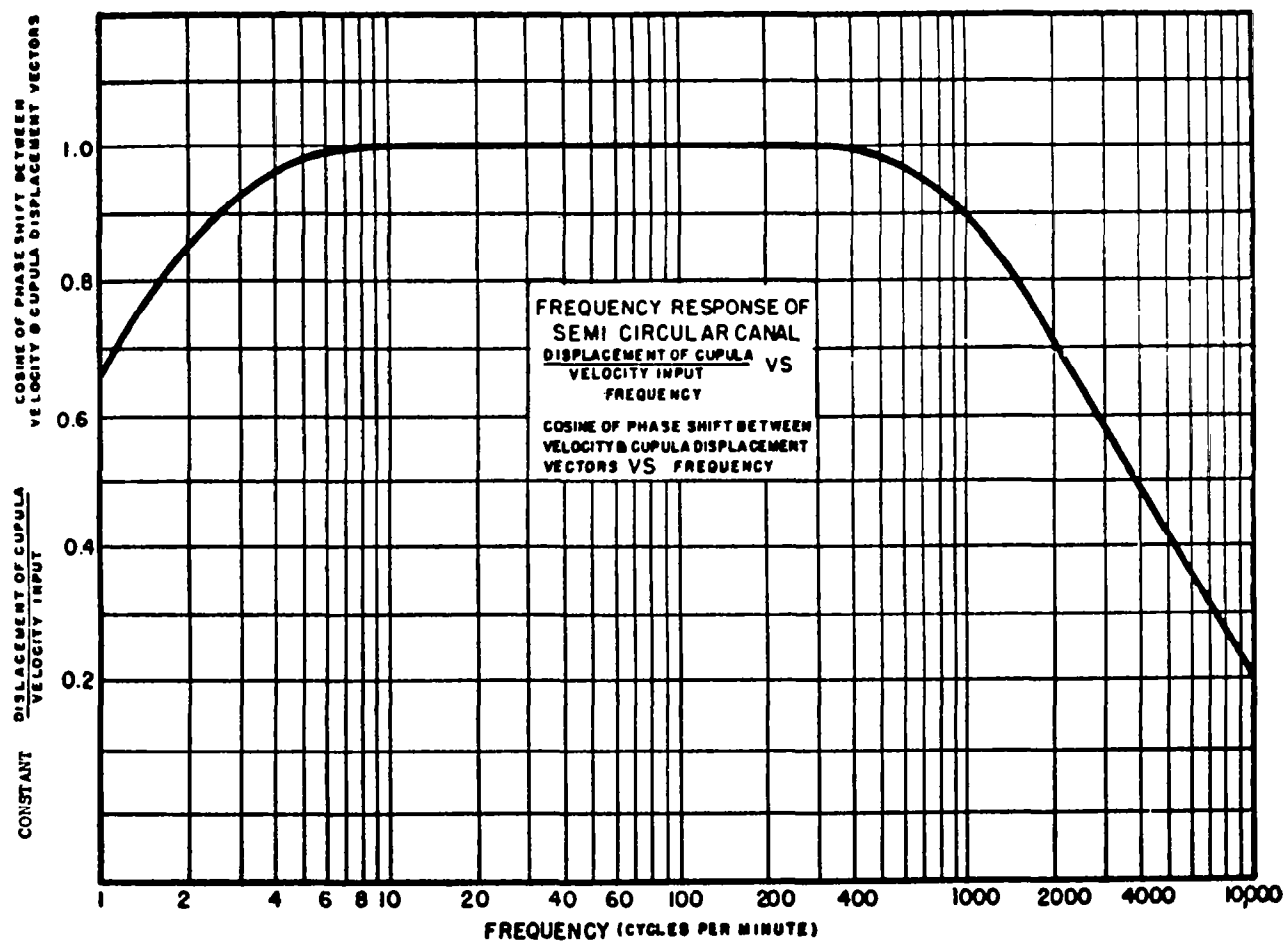


Figure 21. FREQUENCY RESPONSE OF SEMICIRCULAR CANAL (120)

BIBLIOGRAPHY

1. Adams, J.A., L.R. Creamer, Anticipatory timing of continuous and discrete responses. *J. Exper. Psych.*, 63:84-90, 1962.
2. Adrian, E.D. Discharges from vestibular receptors in the cat. *J. Physiol.*, London. 101:389-407, 1943.
3. Agin, L.J., S.G. Marcum, F.J. Daugherty, M.B. Ferrell, and F.N. Trager. Airsickness and related psychosomatic complaints. *Air Surgeon's Bull.*, 2:235, 1945.
4. Armstrong, C.R. Space physiology. *J. Brit. Interplanetary Soc.*, 12:172-175, 1953.
5. Asch, S.E., H.A. Witkin. Studies in space orientation: I. Perception of the upright with displaced visual fields. *J. Exper. Psych.*, 38:325-337, 1948
6. Asch, S.E., H.A. Witkin. Studies in space orientation: II. Perception of the upright with displaced visual fields and with body tilted. *J. Exper. Psych.*, 38:455-477, 1948.
7. Aschan, G. Response to rotary stimuli in fighter pilots. *Acta Oto-laryng.*, Supp. 116, 1954.
8. Bahrick, H.P., P.M. Fitts, G.E. Briggs. Learning curves --facts or artifacts. *Psych. Bull.* 54: 256-268, 1957.
9. Bahrick, H.P., P.M. Fitts, R. Schneider. Reproduction of simple movements as a function of factors influencing proprioceptive feedback. *J. Exper. Psych.*, 49:445-454, 1955.
10. Barlow, J.S. Inertial navigation as a basis for animal navigation. *Neurophysiology Lab. of the Neurol. Service, Mass. General Hospital, Boston, Mass.*
11. Birren, J.E. in *Human Factors in Underseas Warfare*, Chapter 18, National Research Council, Washington D.C., 1949.
12. Bridges, C.C., Bitterman, M.E. The measurement of autokinetic movement. *Amer. J. Psych.*, 67: 525-529, 1954.
13. Brodal, A. Fiber connections of the vestibular nuclei, in *Neural Mechanisms of the Auditory and Vestibular Systems*, Charles C. Thomas, Springfield, ILL., 224-246, 1960.
14. Brodal, A., O. Pompeiano, F. Walberg. The vestibular nuclei and their connections, anatomy and functional correlations. Charles C. Thomas Springfield, Ill., 1962.

15. Brown, J.L., M. Lechner: Acceleration and human performance. J. Av. Med., 27:32-49, 1956.
16. Campbell, E.H. The effectiveness of dramamine in relieving the vestibular reactions following the labyrinthine fenestration operation. Laryngoscope, 59:1261-1272, 1949.
17. Carpenter, M.B. Experimental anatomical-physiological studies of the vestibular nerve and cerebellar connections, in Neural Mechanisms of the Auditory and Vestibular Systems. Ed. G.L. Rasmussen and W.F. Windle, Charles C. Thomas Co., Springfield, Ill., pp297-323.
18. Carmicheal, E.A., M.R. Dix and C.S. Hallpike. Pathology, Symptomatology and diagnosis of organic affections of the eighth nerve system. Brit. Med. Bull., 12:146-152, 1956.
19. Cawthorne, T., et al., The investigation of vestibular function. Brit. Med. Bull., 12:131-142, 1956.
20. Cawthorne, T.E., G. Fitzgerald and S.C. Hallpike. Studies in human vestibular function. II. Observations on the directional preponderance of caloric nystagmus ("Nystagmusbereitschaft") resulting from unilateral labyrinthectomy. Brain 65:138-160, 1942.
21. Chase, R.A., S. Harvey, S. Standfast, I. Rapin and S. Sutton. Studies on sensory feedback. I. Effect of delayed auditory feedback on speech and keytapping. Q.J. Exp. Psych., 13:141-167, 1961.
22. Chinn, H. and P.K. Smith. Motion sickness. Pharmacol. Rev., 7:33-82, 1955.
23. Clark, B. and A. Graybiel. Antecedent visual frame of reference as a contributing factor in the perception of the oculogravic illusion. Bur. Med. & Surg., Project MRO05.13-6001, Subtask 1, Report no. 56.
24. Clark, B. and A. Graybiel. The breakoff phenomenon. J. Av. Med., 28:121-126, 1957.
25. Clark, B. and A. Graybiel. Contributing factors in the perception of the oculogravic illusion. Am. J. Psych., 76:18-27, 1963.
26. Clark, G., and A. Graybiel. Human performance during adaptation to stress in the Pensacola Slow Rotation Room. J. Aerospace Med., 32, 1961.
27. Clark, B. and A. Graybiel. The lag effect associated with stimulation of the semicircular canals as indicated by the oculogyral illusion. U.S. Naval Sch. of Av. Med, Report No. NM 001 059.01.25, 1950.

28. Clark, B., and A. Graybiel. Perception of the postural vertical as a function of practice in normal persons and subjects with labyrinthine defects. Bur. Med. & Surg., Project MR005.13-6001, Subtask 1, Report No. 63, 1961.
29. Clark, B., and A. Graybiel. Vertigo as a cause of pilot error in jet aircraft. J. Av. Med., 28:469-478, 1957.
30. Clark, B. and A. Graybiel. Visual perception of the horizontal following exposure to radial acceleration on a centrifuge. J. Comp. & Physiol. Psych., 44:525-534, 1951.
31. Clark, B., A. Graybiel, and K. MacCorquodale. The illusory perception of movement caused by angular acceleration and by centrifugal force during flight. II. Visually perceived motion and displacement of a fixed target during turns. J. Exp. Psych., 38:298-309, 1948.
32. Clark, B and M.A. Nicholson. Aviator's vertigo: a cause of pilot error in naval aviation students. J. Av. Med., 25:171-179, 1954.
33. Clegg, W.C. and N.M. Dunfield. Non-visual perception of the postural vertical: II. Lateral plane. Canad. J. Psych., 8:80-86, 1954.
34. Collins, W.E. Further studies of the effects of mental set upon vestibular nystagmus. US Army Med. Res. Lab., Fort Knox, Ky., Report No. 443, 1960.
35. Collins, W.E., C.H. Crampton and J.B. Posner. The effect of mental set upon vestibular nystagmus and the electroencephalogram. US Army Med. Res. Lab., Fort Knox, Ky., Report No. 439, 1960.
36. Collins, W.E. and F.E. Guedry, Jr., Arousal effects and nystagmus during prolonged constant angular acceleration. US Army Med. Res. Lab., Fort Knox, Ky., Report No. 500, 1961.
37. Collins, W.E., F.E. Guedry, Jr., and J.B. Posner. Control of caloric nystagmus by manipulating arousal and visual fixation distance. US Army Med. Res. Lab., Fort Knox, Ky., Report No. 485, 1961.
13. Cramer, R.L. Response of mammalian gravity receptors to sustained tilt. J. Aero. Med., 33:663-666, 1962.
39. Crampton, G.H. and W.J. Schwam. Effects of the arousal reaction on nystagmus habituation in cat. US Army Med. Res. Lab., Fort Knox, Ky., Report No. 434, 1960.
40. Crossman, E.R.F.W. The information-capacity of the human motor-system in pursuit tracking. Q.J. Exp. Psych., 12:1-16, 1960.

41. Dearnaley, E.J., et al. Nature and duration of after-sensations following the cessation of turning in a chipmunk aircraft. *J. Aero. Med.*, October, 1962.
42. DeWit, G. Seasickness (motion sickness), a labyrinthological study. *Acta Oto-laryng.*, Supp. 108: 1-56, 1953.
43. Dix, M.R. and C.S. Hallpike. Lesions of the cerebral hemisphere and their effects upon optokinetic and caloric nystagmus. *J. Physiol.*, 114:55, 1951.
44. Dodge, R. Thresholds of rotation. *J. Exp. Psych.*, 6:107-137, 1923.
45. Dohlman, G., Histochemical studies of vestibular mechanisms, in *Neural Mechanisms of the Auditor and Vestibular Systems*. Ed. by G.L. Rasmussen and W.F. Windle, Charles C. Thomas, Springfield, Ill., pp.258-275, 1960.
46. Dohlman, G., J. Farkashidy and F. Saloona. Centrifugal nerve fibers to the sensory epithelium of the vestibular labyrinth. *J. Laryng. Otol.*, 72: 984-991, 1958.
47. Engstrom, H. On the double innervation of the sensory epithelia of the inner ear. *Acta oto-Laryng.*, Stockholm 49:109-118, 1958.
48. Fitts, P.M. The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psychol.*, 47:381-391, 1954.
49. Fitzgerald, G. and C.S. Hallpike. Studies in human vestibular functions: I. Observations on directional preponderance ("Nystagmusbereitschaft") of caloric nystagmus resulting from cerebral lesions. *Brain* 65:115-137, 1942.
50. Fleishman, E.A. The influence of fixed versus free head position on the perception of body position. Human Resources Res. Center, Lackland Air Force Base, Texas, Res. Bull. 53-37, Oct., 1953.
51. Fleishman, E.A. The perception of body position --effect of speed, magnitude and direction of displacement on accuracy of adjustment to an upright position. Human Resources Res. Center, Lackland Air Force Base, Texas, Res. Bull. 53-1 Jan., 1953.
52. Fleishman, E.A. Perception of body position in the absence of visual cues. *J. Exp. Psych.*, 24: 176-187, 1953.

53. Furey, J.A. and R.N. Kraus. A clinical classification of vertigo. *Aeromed. Reviews*, Rev. 7-61, Apr., 1962.
54. Gacek, R.R. Efferent component of the vestibular nerve,,in *Neural Mechanisms of the Auditory and Vestibular Systems*. Ed. by G.L. Rasmussen and W.F. Windle, Charles C. Thomas Co., Springfield, Ill., pp. 276-284, 1960.
55. Gauer, O.H. and G.D. Zuedeman. *Gravitational stress in aerospace medicine*. Little Brown and Co., 1961.
56. Gerathewohl, S.H. and H.D. Stallings. Experiments during weightlessness: A study of the oculo-
agavic illusion. *J. Av. Med.*, 29:504-516, 1958.
57. Gerathewohl, S.J. and H.D. Stallings. The labyrinthine postural reflex (righting reflex) in the cat during weightlessness. *J. Av. Med.*, 28:345, 1957.
58. Gerebtzoff, M.A. Des effets de la stimulation labyrinthique sur l'activite electrique de l'en-
core cerebrale. *C.R. Soc. Biol., Paris*, 131: 807-813, 1939.
59. Gerebtzoff, M.A. Recherches sur la projection corticale de labyrinthe. I. Des effets de la stimulation labyrinthique sur l'activite electrique de l'ecorce cerebrale. *Arch. Int. Physiol.*, 50:59-99, 1940.
60. Gerebtzoff, M.A. Recherches sur la projection corticale de labyrinthe. I. Etude anatomo-
experimentale de la voie vestibulo-cerebrale. *J. Belge Neurol. Psychiat.*, No. 8:417-432, 1940.
61. Gernandt, B. Midbrain activity in response to vestibular stimulation. *Acta Physiol. Scand.*, 21: 73-81, 1950.
62. Gernandt, B.E. and S. Gilman. Generation of labyrinthine impulses, descending vestibular pathways, and modulation of vestibular activity by proprioceptive, cerebellar and reticular influences; in *Neural Mechanisms of the Auditory and Vestibular Systems*. Ed. by G.L. Rasmussen and W.F. Windle, Charles C. Thomas Co., Springfield, Ill., pp. 324-348, 1960.
63. Gooddy, W. and M. Reinhold. Some aspects of human orientation in space. (i) Sensation and movement. *Brain* 75:472-509, 1952.
64. Gooddy, W. and M. Reinhold. Some aspects of human orientation in space. (ii) The dynamic nature of nervous activity. (a)"motion sense" (b) Sense of direction. *Brain* 76:337-363, 1953.

65. Graybiel, A. and R.H. Brown. The delay in visual reorientation following exposure to a change in direction of resultant force on a human centrifuge. *J. Gen. Psych.*, 45:143, 1951.
66. Graybiel, A. and B. Clark. Perception of the horizontal or vertical with head upright, on the side, and inverted under static conditions, and during exposure to centripetal force. *J. Aero. Med.*, 33:147-155, 1962.
67. Graybiel, A., B. Clark and J.J. Zarriello. Observations on human subjects living in a "slow rotation room" for periods of two days. *AMA Arch. Neurol.* 3, July 1960.
68. Graybiel, A., F.E. Guedry, W. Johnson and R. Kennedy. Adaptation to bizarre stimulation of the semicircular canals as indicated by the oculogyral illusion. *J. Aero. Med.*, 32:321-327, 1961.
69. Graybiel, A. and D. Hupp. The oculogyral illusion: A form of apparent motion which may be observed following stimulation of the semicircular canals. *J. Av. Med.*, 17:3, 1946.
70. Graybiel, A. and W.H. Johnson. A comparison of the symptomatology experienced by healthy persons and subjects with loss of labyrinthine function when exposed to unusual patterns of centripetal force in a counter-rotating room. *Ann. Otol. Rhinol. and Laryngol.*, 72:357, 1963.
71. Graybiel, A., W.A. Kerr and S.H. Bartley. Stimulus thresholds of the semicircular canals as a function of angular acceleration. *Am. J. Psych.*, 61:21-36, 1948.
72. Graybiel, A., J.I. Niven, T.E. Walsh. The differentiation between symptoms referable to the otolith organs and semicircular canals in patients with nonsuppurative labyrinthitis. *Laryng.*, 62:924-933, 1952.
73. Groen, J.J., O. Lowenstein and A.J.M. Vendrik. The mechanical analysis of the responses from the end organs of the horizontal semicircular canal in the isolated elasmobranch labyrinth. *J. Physiol.*, London, 117:329-346, 1952.
74. Guedry, F.E., Jr. and N. Beberman. Apparent adaptation effects in vestibular reactions. *US Army Med. Res. Lab.*, Fort Knox, Ky., Report No. 293, 1957.
75. Guedry, F.E., Jr. and S.J. Ceran. Derivation of 'subjective velocity' from angular displacement estimates made during prolonged angular accelerations: adaptation effects. *US Army Med. Res. Lab.*, Fort Knox, Ky., Report No. 376, 1959.

76. Guedry, F.E., Jr., W.E. Collins, and P.L. Sheffey. Perceptual and oculomotor reactions to interacting visual and vestibular stimulation. US Army Med. Res. Lab., Fort Knox, Ky., Report No. 463, 1961.
77. Guedry, F.E., Jr., R.L. Cramer and W.P. Koella. Experiments on the rate of development and rate of recovery of apparent adaptation effects in the vestibular system. US Army Med. Res. Lab., Fort Knox, Ky., Report no. 336, 1958.
78. Guedry, F.E., Jr. and A. Graybiel. The appearance of compensatory nystagmus in human subjects as a conditioned response during adaptation to a continuously rotating environment. Bur. Med. & Surg., Project MRO05.13-6001, Subtask 1, Report No. 61.
79. Guedry, F.E., Jr. and A. Graybiel. Compensatory nystagmus conditioned during adaptation to living in a rotating room. J. App. Physiol., 17, May 1962.
80. Guedry, F.E., Jr. and H. Kalter. Description of human rotation device. US Army Med. Res. Lab., Fort Knox, Ky., Report No. 242, 1956.
81. Guedry, F.E., Jr., and L.S. Lauver. The oculomotor and subjective aspect of the vestibular reaction during prolonged constant angular acceleration. US Army Med. Res. Lab., Fort Knox, Ky., Report No. 438, 1960.
82. Guedry, G.E., Jr. and E.K. Montague. Relationship between magnitudes of vestibular reactions and effective coriolis couples in the semicircular canal system. US Army Med. Res. Lab., Fort Knox, Ky., Report no. 456, 1960.
83. Guedry, F.E., Jr., L.J. Peacock and R.L. Cramer. Nystagmic eye movements during interacting vestibular stimuli. US Army Med. Res. Lab., Fort Knox, Ky., Report no. 275, 1957.
84. Hammer, L.R. Perception of the visual vertical under reduced gravity. Tech. Doc. Report no. MRL-TRL-62-55, Wright Patterson AFB, Ohio, May 1962.
85. Hartinger, H., Motion and perception of space. USAF Sch. Av. Med., Special Report, 1951.
86. Hartman, B.O. The effect of target frequency on compensatory tracking. US Army Med. Res. Lab., Fort Knox, Ky., Report no. 272, 1957.

87. Hawkes, G.R., R.W. Bailey and J.S Warm. Method and modality in judgements of brief stimulus duration. US Army Med. Res. Lab., Fort Knox, Ky., Report no. 442, 1960.
88. Hemingway, A. Results of 500 swing tests for investigating motion sickness. AAF Sch. Av. Med., Randolph AFB, Res. Project 31, Report no. 2, 1942.
89. Henry, F.M. Dynamic kinesthetic perception and adjustment. Res. Q. Am. Assoc Health & Phys. Ed., 24:176-187, 1953.
90. Henry, J.P., et al. Effects of weightlessness in ballistic and orbital flight. J. Aero. Med., Sept., 1962.
91. Hick, W.E. The discontinuous functioning of the human operator in pursuit tasks. Q.J. Exp. Psych., 1: 36-52, 1948.
92. Hixson, W.C. and J.I. Niven. Applications of the system transfer function concept to a mathematical description of the labyrinth: I. Steady-state nystagmus response to semicircular canal stimulation by angular acceleration. Bur. Med. & Surg., Project MRO05.13-6001, Subtask 1, Report No. 57, 1961.
93. Howland, D. and M.E. Noble. The effect of physical constants of a control on tracking performance. J. Exp. Psych., 46:353-360, 1953.
94. Johnson, W.H. Head movement measurements in relation to spatial disorientation and vestibular stimulation. J. Av. Med., 27:148-152, 1956.
95. Johnson, W.H. The importance of head movements in the production of motion sickness. J. Def. Res. Bd., Canada, Library of Congress, Tech. Info. Div. No. U22185, Apr. 1952.
96. Johnson, W.H. and J.W. Mayne. Stimulus required to produce motion sickness: Restriction of head movements as a preventive of airsickness --field studies on airborne troops. J. Av. Med., 24: 400-411, 1953.
97. Johnson, W.H., J. Meek and A. Graybiel. Effects of labyrinthectomy on canal sickness in squirrel monkey. Ann. Otol-Rhino-Laryngo., 71:289, 1962.
98. Johnson, W.H., J. Meek and A. Graybiel. The effects of unilateral and bilateral labyrinthectomy on canal sickness in the squirrel monkey. Bur. Med. & Surg., Project MRO05.13-6001, Subtask 1, Report No. 65.
99. Johnson, W.H. and B. Smith. Functions of the utricular macula. A preliminary report. Can. J. Biochem. Physiol., 37:605-606, 1959.

100. Kennedy, R.S. and A. Graybiel. Symptomatology during prolonged exposure in a constantly rotating environment at a velocity of one revolution per minute. Bur. Med. & Surg., Project MR005.13-6001, Subtask 1, Report no. 62.
101. Kennedy, R.S. and A. Graybiel. Symptomatology during prolonged exposure in a constantly rotating environment at a velocity of one revolution per minute. J. Aero. Med., 33:817-825, 1962.
102. Kraus, R.N. Disorientation in flight: an evaluation of the etiological factors. Aero. Med., 30:664-673, 1959.
103. Lansburg, M.P. Some considerations and investigations in the field of labyrinthine functioning. IV. Influences of vestibular-stimulation upon electroencephalogram, electrocardiogram and respiration. Aeromedica Acta, (Soesterberg, Netherlands) 3:241-246, 1954.
104. Levy, C.K. and H. Quastler. Acute responses of the vestibular apparatus to X-irradiation. Rad. Res., 16:189-200, 1962.
105. Loftus, J.P., ed. Symposium on motion sickness with special reference to weightlessness. Tech. Doc. Rep. No. AMRL-TDR-63-25, June 1963.
106. Lowenback, H. and W.H. Gantt. Conditional vestibular reactions. J. Neuro. Physiol., pg. 4348, 1940.
107. Lowenstein, O. The equilibrium function of the vertebrate labyrinth. Biol. Rev., 11:113-145, 1936.
108. Lowenstein, O. and A. Sand. The activity of the horizontal semicircular canal of the dogfish, *Scyllium canicula*. J. Exp. Biol., 13:416-428, 1936.
109. Lowry, R.H. and W.H. Johnson. 'Pseudo motion sickness' due to sudden negative "G": its relation to 'airsickness'. J.Av. Med., April, 1954.
110. MacCorquodale, K. Effects of angular acceleration and centrifugal force on nonvisual space orientation during flight. J. Av. Med., 19: 146-157, 1948.
111. Magnus, R. Animal posture (The Croonian lecture). Proc. R. Soc. (B.), 98:339-353, 1925.
112. Magnus, R. Some results of studies on the physiology of posture. Lancet, pp. 531-536 and 585-588, 1926.
113. Mann, C.W. Final technical report. Joint project report No. 43, USN Sch. Av. Med., Pensacola, Florida, 1956.

114. Mann, C.W. Visual factors in the perception of verticality. J. Exp. Psych., 44:460-464, 1952.
115. Mann, C.W. Which way is up? Res. Rev., ONR, Dept. of the Navy. October 1951.
116. Mann, C.W., N.H. Brrthelot-Berry and H.J. Dauterive, Jr. The perception of the vertical: I. Visual and non-labyrinthine cues. J. Exp. Psych., 39: 700-707, 1949.
117. Mann, C.W., C.J. Canella and A. Graybiel. An examination of the technique of cupulometry. Joint project report No. 42, USN Sch. Av. Med., Pensacola, Florida, May 1956.
118. Mann, C.W. and G.E. Passey. The perception of the vertical: V. Adjustment to the postural vertical as a function of postural tilt and duration of exposure. J. Exp. Psych., 41:108-113, 1951.
119. Mann, C.W. and J.T. Ray. The perception of the vertical: XIV. The effect of the rate of movement on the judgement of the vertical. USN Sch. Av. Med. and Tulane U., Joint project report No. 40, 1956.
120. Mayne, R. Some engineering aspects of the mechanism of body control. Electrical Engineering 70: 207-212, 1951.
121. McCabe, B.F. Vestibular suppression in figure skaters. Trans. Amer. Acad. Ophth. and Otol., 64:1960.
122. McFarland, R.A. Human factors in air transportation. McGraw-Hill, N.Y., 1953.
123. McNally, W.J. The otoliths and the part they play in man. Laryngoscope, 54:304-323, 1944.
124. McNally, W.J. and E.A. Stuart. Physiology of the labyrinth reviewed in relation to seasickness and other forms of motion sickness. War Med., 2:683-771, 1942.
125. Mickle, W.A. and H.W. Ades. Rostral projection pathways of the vestibular system. Am. J. Physiol., 176:243-252, 1954.
126. Miller, E.F. and A. Graybiel. A comparison of ocular counterrolling movements between normal persons and deaf subjects with bilateral labyrinthine defects. Joint Report, Bur. Med. & Surg., Project MRO05.13-6001, Subtask 1, Report No. 68, 1962.

127. Miller, E.F. and A. Graybiel. Role of the otolith organs in the perception of horizontality. Joint Report, Bureau Med. & Surg., Project MR005.13-6001, Subtask 1, Report No. 80, 1963.
128. Moore, E.W. and R.L. Cramer. Perception of postural verticality. USAF Sch. Av.Med., TDR-62-72, June 1962.
129. Morant, R.B. The visual perception of the median plane as influenced by labyrinthian stimulation. J.Psychol. 47:25-35, 1959.
130. Morrell, F. Electrophysiological contributions to the neural basis of learning. Physiol. Rev. 44:3-494, 1961.
131. Niven, J.L. and W.C. Hixson. Frequency response of the human semicircular canals: steady-state ocular nystagmus response to high-level sinusoidal angular rotations. Bureau Med.& Surg., Project MR005.13-6001, Subtask 1, Report No. 58.
132. North, J.D., Z.A.Lomnicki and S.K.Zaremba. The design and interpretation of human control experiments. Ergonomics. 1:314-327, 1957-58.
133. Nuttall, J.B. The problem of spatial disorientation. J.A.M.A. 166:431-438, 1958.
134. Nuttall, J.B. and W.G. Sanfor. Spatial disorientation in operational flying. USAF Norton AFB, San Bernardino, Calif., Report No. M-27-56, May 1956.
135. Passey, G.E. The perception of the vertical. IV. Adjustment to the vertical with normal and tilted visual frames of reference. J.Exp. Psychol. 40:738-745, 1950.
136. Passey, G.E. and F.E. Guedry, Jr. The perception of the vertical: II. Adaptation effects in four planes. J.Exp.Psychol. 39:700-707, 1949.
137. Petroff, A.E. An experimental investigation of the origin of efferent fiber projections to the vestibular neuro-epithelium. (Abstract) Anat. Rec. 121:352-353, 1955.
138. Pfaltz, C.R. Die Normale Calorische Labyrinthreaktion. Arch. Ohren- usw. Heilk. u.Z.Hals- usw. Heilk. 172:131-174, 1957.
139. Pfaltz, C.R. and R. Gulick. Die Pathologische Calorische Labyrinthreaktion. Arch. Ohren- usw. Heilk. u.Z. Hals- usw. Heilk. 179:525-544, 1962.
140. Pfaltz, C.R. and R.R. Pfaltz. Zur Prufung der Wirksamkeit Vestibular Angreifender Pharmaka. HNO 11:165-169, 1963.

141. Pigg, L.O. and Kama, W.N. Visual Acuity in relation to body orientation. Wright-Patterson AFB, Tech. Doc. Report MRL TDR-62-74, 1962.
142. Poulton, E.C. On the stimulus and response in pursuit tracking. J.Exp. Psychol. 53:189-194, 1957.
143. Quix, F.H. Epreuves Vestibulaires. L'Oto-Rhino-Laryng. Internat. 6:484, 1922.
144. Rasmussen, G.L. And R. Gacek. Concerning the question of an efferent fiber component of the vestibular nerve of the cat. (Abstract) Anat. Rec. 130:361-362, 1958.
145. Razran, G. The observable unconscious and the inferable conscious in current soviet psychophysiology: Interoceptive conditioning, semantic conditioning and the orienting reflex. Psych. Rev. 68:1961.
146. Richter, Hs.R. and C.-R. Pfaltz. Recherches nystagmo-graphiques sur la reaction vestibulaire galvanique. Confinia Neurologica 18:370-377, 1958.
147. Roman, J.A., B.H. Warren and A. Graybiel. Observation of the elevator illusion during subgravity preceded by negative accelerations. Joint Report, Bureau Med.&Surg., Project MR005.13-6001, Subtask 1, Report No. 83, 1963.
148. Roman, J.A., B.H. Warren and A. Graybiel. The sensitivity to stimulation of the semicircular canals during weightlessness. Joint Report, Bureau Med. & Surg., Project MR005.13-6001, Subtask 1, Report No. 84, 1963.
149. Scher, S.H. Pilot's loss of orientation in inverted spins. National Advisory Committee for Aeronautics, Tech. Note 3531, Langley Aeron.Lab., Langley Field, Va., Oct. 1955.
150. Schock, G.J.D. Apparent motion of a fixed luminous target during sub-gravity trajectories. Tech. Report 58-3, AF Missile Development Center, Holloman AFB, N.M., 1958.
151. Schwab, R.S. The nonlabyrinthine causes of motion sickness. Internat. Record Med., 167:631-637, 1954.
152. Simons, D.G. Sub-gravity and weightlessness in gravitational stress in aerospace medicine. Ed. by Gauer and Zuidema, Little Brown and Co., Boston, 189-201, 1961.
153. Smith, P.E. On "Physiological Responses" in Simulation Council Newsletter, ed. J.McLeod, p.119 Aug 1962, in Instruments and Control Systems.

154. Solley, C.M. Reduction of error with practice in perception of the postural vertical. J.Exp.Psychol. 52:329-333, 1956.
155. Spiegel, E.A. Effect of labyrinthine reflexes on the vegetative nervous system. Arch.Oto-Laryng. 44: 61, 1946.
156. Spiegel, E.A., G.C. Henry and H.T. Wycis. Changes of cerebral circulation induced by labyrinthine stimulation. Am.J. Physiol., 142:589-593, 1944.
157. Spiegel, E.A., M.J. Oppenheimer, G.C. Henry and H.T.Wycis. Experimental production of motion sickness. War Med., 5:283, 1944.
158. Spiegel, E.A. and J. Sommer. Vestibular Mechanisms. Med. Physics. ed by O.Glasser. The Year Book Publishers inc., Chicago, p.1638-1953, Vol. I and p.399-401, Vol II, 1950.
159. Steele, J.E. Motion sickness and spatial perception: A theoretical study. ASD Technical Report 61-530, 1961.
160. Steinhausen, W. Über den Nachweis der Bewegung der Cupula in der Intakten Bogengangsampeulle des Labyrinthes bei der Naturlichen Rotatorischen und Calorischen Reizung. Pflug. Arch. ges. Physiol. 228:322-328, 1931.
161. Tyler, D.B. and P.Bard. Motion Sickness. Physiol. Rev. 29: 311-369, 1949.
162. Useller, J.W. and J.S. Algranti. Pilot control of space vehicle tumbling. Nat. Specialists Meeting on Guidance of Aerospace Vehicles, N.Y. 1960, pp.109-113, 1960.
163. Useller, J.W. and J.S. Algranti. Pilot reaction to high-speed rotation. IAF Meeting, Stockholm, 1960.
164. U.S. Dept. H.Ed.& W. Public Health Service, National Library of Medicine. Ref. Div., Bibliography of Space Medicine. Publication No. 617, 1958.
165. Van Egmond, A.A.J., J.J. Groen and L.B.W. Jongkees. The mechanics of the semicircular canal. J. Physiol. 110:1-17, 1949.
166. Van Sickle, N.D. Modern Airmanship, 2nd ed. New York, D. Van Nostrand & Co. Inc., 1961.
167. Vinacke, W.E. Aviator's vertigo. J.Av.Med. 19:158-170, 1948.
168. Von Békésy, G. Neural volleys and the similarity between some sensations produced by tones and by skin vibrations. J. Acoustical Soc. Am. 29:1059-1070, 1957.

169. Von Holst, E. and H. Mittel staedt. Das Reafferenzprinzip. (Wechselsirkungen zwischen Zentralnervensystem un Peripherie). Z.Naturw. 37:464, 1950.
170. Walsh, E.G. The perception of rhythmically repeated linear motion in the horizontal plane. Brit. J. Psychol. 53:439-445, 1962.
171. Wapner, S. and H.A. Witkin. The role of visual factors in the maintenance of body-balance. AM.J. Psychol. 385-408, 1950.
172. Webb Associates, NASA Life Sciences Data Book. Director of Aerospace Medicine, Contract R-89, NASA, 1962.
173. White, W.J. Acceleration and vision. USAFWADC TR 58-333, Aero-Medical Laboratory, Nov. 1958.
174. Whitsett, C.E., Jr. Some dynamic response characteristics of weightless man. Tech. Doc. Report No. AMRL-TDR-63-18, Wright-Patterson Air Force Base, Ohio, Apr. 1963.
175. Witkin, H.A. Further studies of perception of the upright when the direction of the force acting on the body is changed. J.Exp.Psychol. 43:9-20, 1952.
176. Witkin, H.A. The nature and importance of individual differences in perception. J. Personality 18:145-170, 1949.
177. Witkin, H.A. and S.E. Asch. Studies in Space Orientation. III. Perception of the upright in the absence of a visual field. J.Exp.Psychol. 38:603-614, 1948.
178. Witkin, H.A. And S.E. Asch. Studies in space orientation. IV. Further experiments on perception of the upright with displaced visual fields. J.Exp.Psychol. 38:762-782, 1948.
179. Witkin, H.A. and S. Wapner. Visual factors in the maintenance of upright posture. Am.J. Psychol. 63:31-50, 1950.
180. Woellner, R.C. The perception of vertical in the presence of increased accelerative forces. Bur. Med. & Surg., Research Project NM 17 01 11, Subtask 1, Report No. 45, 1957.
181. Woellner, R.C. and A. Graybiel. Counterrolling of the eyes and its dependence on the magnitude of gravitational or inertial force acting laterally on the body. J. App. Physiol. 14:632-634, 1959.
182. Woellner, R.C. and A. Graybiel. The loss of counterrolling of the eyes in three persons presumably without functional otolith organs. Ann. Otol., Rhin. & Laryng. 69:1006-1012, 1960.
183. Worchel, P. The role of the vestibular organs in space orientation. J. Exp. Psychol. 44:4-10, 1952.

184. Worchel, P. The vestibular organs in space orientation. Perceptual and Motor Skills 5:164, 1955.
185. Ziebolz, H. and H.M. Paynter, Possibilities of a two-time scale computing system for control and simulation of dynamic systems. Proc. Nat. Electronic Conf. Vol. 9, February 1954.
186. Ziegenruecker, G.H. And E.B. Magid. Short time human tolerance to sinusoidal vibrations. USAF WADC TR 59-391, Aero-Medical Laboratory, July 1959.

Additional References:

187. Fogel, Lawrence J., Biotechnology: concepts and applications. "The Position- and Motion-Sensing Channel" pp. 145-165, Prentice Hall, Inc. Englewood Cliffs, N.H., 1963.
188. Clark, B. and A. Graybiel. Visual perception of the horizontal during prolonged exposure to radial acceleration on a centrifuge. J.Exp.Psychol. 63: 294-301, 1962.

SECTION 3

PUPIL AND LENS CONTROL SYSTEMS

by Lawrence Stark

Summary

In the last decade the experimental approach to biocontrol systems has justified the hopeful observations of early cyberneticians. The development of systems approaches to control and communication in the engineering world has provided the analytical biologist with concepts, analogs, analysis methods, and necessary instrumentation. Initially, linear mathematical concepts and analysis methods were used as transfer function descriptions to shed light on such characteristic properties of neurological feedback control systems as stability, oscillations and noise. Further studies penetrated essential nonlinear, discontinuous and adaptive properties of neurological servomechanisms and revealed two broad classes of these biological feedback control systems with quite different properties.

The first class, including eye and hand target tracking has intermittent, input adaptive characteristics which dominate their behavior. The input-synchronized intermittency operator shows up as refractoriness when two sequential responses are required, or as sampled data peaks in Bode plots. These systems appear fairly linear, that is twice the input yields twice the output. On closer examination, however, certain of their trajectories can be approximated by

nonlinear optimal control analogs. The ability to predict repetitive characteristics of the input signal make them highly input-adaptive. It is necessary to control the effect of this powerful prediction operator by using stochastic inputs in experiments designed to elucidate underlying properties. Sampled data models for hand and eye movements have had considerable success in consolidating many phenomena associated with these neurological servomechanisms and in predicting behavior under new circumstances.

The second class, including the pupil and temperature control, is highly nonlinear and generally show biological adaptation. By this we mean that the reference input adjusts to accept new environmental conditions. For example, retinal adaptation permits the pupil to accept the recent-past illumination level as the reference, and thus regulation occurs with respect to this shifting or adapting reference. Multiple scale compression nonlinearities provide for operating over many log units of input range. Separate mechanisms are used in different time and frequency ranges. These may have different gains so that asymmetrical responses are prominent. Limit cycles appear and can be studied using Wiener's G-functional analysis to canonically describe input-output characteristics of these systems, but simpler more classical methods such as describing-function techniques are often revealing.

Cybernetics as a sound scientific approach rests heavily upon mathematics as a language. Superficial verbal aspects do not demonstrate the beauty and power of this type of research. We have, therefore, decided to select two systems, the lens and the pupil for further discussion in this section. With so narrow a field we may review experimental and analytical results which exemplify the essential nature of these bioengineering studies.

Now that biochemistry and biophysics have split off from physiology, that subject centers upon system organizations and systems function. The chemical and physical componentry may be quite different from the engineering components in space craft and computers but the analogies between engineering and biology at the systems level is close indeed. It is hoped that this review will demonstrate the effectiveness of cybernetics, in particular the application of feedback control theory, in theoretical and experimental analysis of neurophysiological servomechanisms.

In our contract for a review of the field of biological control systems, previous reports have dealt with introductory material on control theory, the vestibular system, the manual control system, biochemical and hormonal control systems, body fluid control system, and control of sensation from the skin.

As part of the overall study, this report concentrates on two important aspects of the visual motor control system. These aspects pertain to the formation of sharp images of appropriate light intensity on the retina rather than the positioning of these images. The servomechanisms which control the pupil diameter and lens accommodation are evaluated and the current state of knowledge reviewed. The visual motor control systems pertaining to positioning of the image, namely, control of eye movements, will be considered in a future report.

I. PUPIL

A. Input-output identification.

The iris muscles, surrounding the pupil or aperture of the eye, act to regulate light impinging upon the retina. In addition to this primary light regulating function changes of pupillary area occur in synkinesis with lens accommodation and with vergence movements of the eyeballs as indicated in Fig. 1.

When pupillary aperture is small less light flux enters the eye, only the central part of the lens is utilized, and the depth of focus of the eye is greater. The variation of pupil area, or rms noise increases also. When the pupil area is large more light flux can enter the eye, the periphery of the lens with greater aberrations is used, depth of focus

diminishes, and pupil noise becomes lessened.

B. Element anatomy and physiology from an engineering point of view.

The pupil is so widely observed an organ that most persons are already acquainted with certain basic facts of its anatomy and physiology. The pupil is the hole in the center of the iris muscle which enables light to enter the eye and impinge upon the retina, the sensitive layer of the back of the eye. The retina is comprised of primary sensory cells containing photosensitive pigments which trap photons and subsequently stimulate retinal nerve cells. The retina is part of the central nervous system and possesses a complex multineural integrative (i.e. information transforming) apparatus. The optic nerve leads mainly to the visual cortex of the cerebral hemispheres via a relaying station, the lateral geniculate body. However, some fibers called the pupillomotor fibers, go directly to the brain stem and relay in the pretectal area and thence to the Edinger-Westphal nucleus. This nucleus contains the nerve cells, part of the parasympathetic system, whose fibers (after an external relay in the ciliary ganglion) control the powerful sphincter muscle of the iris. Fiber tracts also go to the sympathetic system in the spinal cord. Here, nerve cells send fibers back to the orbit, after relaying in the superior cervical ganglion. The dilator of the pupil is controlled by these sympathetic

fibers and is, for example, responsible for the wide dilatation of the pupils after local administration of adrenalin (30).

Excitation of the Edinger-Westphal nucleus produces constriction of the pupil and it is also probable that inhibition, i.e., decrease in the operating level, of this nucleus is also the most important mechanism for dilating the pupil (30).

Figure 2 (23) shows a mechanical diagram representation of the sphincter and dilator muscles of the iris. This complex structure still requires much further study. For example Apter (3) denies the ability of the dilator to actively contract while Simpson (25) has experimentally demonstrated apparent dilator contraction in man. This neuromuscular mechanism is sensitive to many drugs, a result of its complex double autonomous nervous system innervation and a partial list of such effective pharmacological agents is also presented in Figure 2.

The neuroanatomy of the pupillary system is also not completely understood. It is believed that accommodation synkinetic signals join the light signals at the final common pathway, the Edinger-Westphal nucleus, of the parasympathetic outflow. The retina with its many layers of neurons relays to the even more complex visual cortex and essentially no knowledge of the exact functions of these most important

neural suborgans exists. By analogy to the work of Lettvin and McCulloch (19) on the frog retina and to work on mammalian visual cortex (17) it would appear that complex operators are acting on the information originally contained in the image. The mean light flux is among the least important of image parameters and it may be that only the brainstem directly relays this via the pretectal nucleus to the Edinger-Westphal nucleus as suggested in the neuroanatomical review by Magoun and Ranson, 1935 (22). However, complex changes in visual sensitivity secondary to saccadic eye movement also control pupil sensitivity and this suggests more indirect pathways. Both rods and cones play a role in pupillary reactions as indicated by the finding of Abelsdorff (1) that the Purkinje shift of visual spectral sensitivity with dark adaptation also occurred in pupil sensitivity.

The origin of the accommodation and vergence synkinetic pathways is conjectured to be in the visual cortex and Jampel has demonstrated areas in monkey brain which when stimulated produce these reactions. (18)

C. Older physiological behavioral studies.

A review giving the flavor of older physiological behavioral studies is "The Pupil", Chapter 9 in Volume 3 of The Eye, edited by Hugh Davson (21). This work concerns itself especially with the diagnostic possibilities of pupillometry. It is clear that as many and various elements

interact in producing the complex behavior of the pupillary system it is no easy task to uniquely define defective operation of a particular element from black-box observations of pupillary reaction shapes.

Figure 3 (29) shows typical behavior of the pupil with temporary constrictions of pupil area occurring in response to light flashes. Of interest also is the long lasting pupillary constriction occurring synkinetically with voluntary accommodation. As will be shown later in the discussion of control systems experiments and models the dependence of amplitude of response upon mean area is a strong nonlinearity. Also visible in Figure 3 is the increase of area variability or noise as area itself decreases.

Lowenstein and Givner (20) showed that short periods of decreased light intensity or of darkness also produces pupillary constriction, a result of asymmetry of rates of constriction and dilation apparent in part in such large-signal studies of the pupil as those of Young and Biersdorf (40) and in small-signal studies by Clynes (10) and others (27).

Of special interest has been a large number of attempts to obtain pupillary conditional reflex responses. This would seem to be possible via the multiple involvement of pupillary dilatation in alerting or ortho-sympathetic responses or via the semi-voluntary accommodative synkinetic pupillary response.

However, although many early studies attempted and reported positive results in classical conditioning paradigms with the pupil reflex response to light as the unconditioned stimulus-response pair, Young (39) clearly showed the inadequacy of non-objective measurements in his 1954 paper, and no subsequent demonstrations have appeared.

D. Control systems experiments.

The classical paper in 1957 (31) experimentally measuring the open-loop transfer function of the pupil light reflex initiated a series of studies on this and other systems (34). Figure 4 (30) illustrates the Bode diagram of the pupil system which is approximated by the transfer function:

$$G(s) = \frac{0.16e^{-0.2s}}{(1 + 0.1s)^3} \quad (1)$$

The pupil reflex to light has been considered as a servomechanism, a self-regulated error actuated control device. This cybernetic approach, requiring the experimenter to make quantitative measurements in animals with a fully intact central nervous system, was made possible using a pupillometer designed for awake, cooperative human subjects. This instrument provided an electronically controlled light stimulus as well as continuous records of both pupil area and light intensity. Sinusoidal changes in light intensity,

small enough for linearization assumptions, were injected in an open loop fashion to determine the transfer function for pupil system behavior. The pupil servo is quite stable and has a low gain with an attenuation slope of 18 db per octave beyond 1.5 cps. One line of investigation using pharmacological agents has suggested the triple lag to be contributed by the physical law representing the viscosity of the iris neuromuscular system. Another experiment used artificially increased gain to produce instability oscillations, whose frequency was predictable from the low gain transfer function. Still another investigation has shown the pupil system to contain much noise. This noise is not a result of instability, nor generated by the smooth muscle of the iris, but is probably produced as wide band flicker noise (?due to asynchrony in nerve impulse firings) in the Edinger-Westphal nucleus, and frequency shaped by output elements. Further studies in progress are defining nonlinearities in the pupil and retinal system in order to set up an accurate hybrid analog digital computer model of the pupil system. The general manner in which pupil behavior is shaped for various operating ranges and frequencies throws light on the power of an adaptive servomechanism to maximize its utility to the organism. The value of the cybernetic approach is demonstrated by both the clarification which these concepts, derived from control and communications engineering, have introduced into the understanding of pupil behavior as well as by the precision of the

experimental data obtained using measurement techniques adapted for physiological purposes. (30)

Since these early nonlinear studies a good deal of further work has ensued. Describing function techniques have not proven as useful as heuristic experimentation in conjunction with the analog-digital computer modelling system. Other possible more powerful approaches will be discussed in section F.

E. Models

A model provides for a concise summary of the system of complexly interacting nonlinear functional relations that have been discovered, analyzed, and studied in a variety of experiments. As such, models from the early transfer function of 1957 to the Wiener nonlinear kernels of 1965 have played an important role in the cybernetic analysis of the pupillary system.

One such model (26) is shown in Fig. 5. Light enters as L and is operated upon by a lead-lag retinal adaptation element with parameters as given. The signal, V , then goes into a saturating nonlinearity whose functional form may be given by a table lookup or computed as

$$V = \log (1 + u)$$

The V signal now traverses two paths; an upper 1-G path which is in operation at all times, and a lower path which is open only for positive changes of V . This lower path has an additional lag element and is further multiplied by a white Gaussian noise source. This provides for noise as both a stochastic function of $(1 - \bar{A})$ and as a lagged function of $(1 - A)$ when the noise is determined as an instantaneous function of A by ensemble averaging.

The two paths join as signal, P , at the summer. The upper path is open only for positive changes of P and the lower path is then closed. With negative changes of P a long time constant governing dilatation operates and is noise free. After the second summer the signal Q is operated on by the transport delay and the triple lag. Then an inversion with respect to A_{\max} occurs to produce the output, A . The retina, central nervous system and pupil muscle dynamics portion of the system have been determined by multi input - multi output black-box experiments, by pharmacological and surgical dissection experiments, and by use of older anatomical and physiological experimental results. Other models have been developed in more detail for various specific portions of the entire system but less completely for the system as a whole.

F. Critical Summary

Before finally summarizing the state of knowledge of the pupillary system a brief discussion of some newer methods is relevant. Firstly the attractiveness of applying a canonical approach to nonlinear systems has led to the graphs of Fig. 6 (28). This shows $h_1(T)$ the first order kernel and $h_2(T_1, T_2)$ the second order kernel as two and three dimensional functions respectively. The higher order G-functional analysis of Wiener (38) provides a statistical bound (much like thermodynamics) to the behavior of systems. The curve above the 45° diagonal represents no-memory nonlinearities of second-order while the remainder of the curve (that is statistically significant) represents memory-dependent nonlinearities to second-order.

In addition to this mathematical approach a good deal more work is required in terms of input-output analysis of the pupil in conjunction with heuristic hybrid analog-digital computer models and further exploitation of the multi-input (two eyes, accommodation, memory, fatigue) and multi-output (two eyes, noise) analytical approach. The use of conditioned animals and patients with various neurological syndromes should also provide more quantitative data and new ideas about locations of nonlinear and linear operators.

Noise studies suggest asynchronous neural firing as the shot or flicker noise source that would be equivalent to the multiplicative noise in the system model. Further experiments with implanted electrodes and with isolated neuromuscular preparations will be helpful. In short we require more experimentation on the elements and the whole systems and more analysis both statistical and heuristic.

However, a good start on this complex system has already been made and our models are excellent predictors of most of the obvious experimental results, and suggestive of further crucial experiments. They are especially accurate in the crucial control areas relating to stability and the generation of unstable oscillatory behavior.

II. LENS

A. Input-Output Identification

A target moving closer and farther from an eye with a fixed accommodative state would produce a retinal image that develops more or less "blur" as the target went further or closer to the particular "clear vision position" at which accommodation is fixed. The accommodative mechanism of the normal human eye has a variable power lens which by changing can alter the clear vision position and so attempt to keep sharp the retinal image of the target.

Thus target distance input and clear vision position as output are defined physical quantities. The difference between these two, or distance error, represents a classical servoanalytical error signal which may serve to drive the accommodative mechanism. The optical apparatus of the eye, however, transforms the distance error to an amount of blur or sharpness error which is again a well defined physical quantity, a distribution of light on the retinal surface which might, for example, be photographed (alas, through the dioptric apparatus of the eye). These variables are indicated in the block diagrams of Fig. 7a and 7b (33).

Several synkinetic or associated movements of related systems also appear as outputs in Fig. 7a. These are the pupillary and the accommodative-vergence responses. If

the target moves on the optical axis of the seeing eye under monocular viewing conditions, the nonseeing eye (viewing only an unstructured field, for example) will change its vergence posture and incidentally indicate the state of accommodation. Similarly the pupil constricts with near targets and dilates for distant targets, changing both light intensity and depth of focus.

B. Element Anatomy and Physiology from an Engineering Point of View

As can be seen in Fig. 7b there are four blocks of elements in the lens system.

1. Lens optics was intensively studied by Thomas Young, who in 1801 showed that changes in shape of the lens must be responsible for changes in accommodation and not, for example, changes in corneal curvature or changes in length of the eyeball (See Fig. 8)(11). Fincham (1937) has produced quantitative evidence concerning the variable bulging of the central portion of the anterior surface of the lens where the lens capsule is thinnest and the radius of curvature least (13).

2. The target image is sensed by the photoreceptors of the retina and the light distribution represents a greater or lesser amount of blur according as the distance error is greater or lesser. While distance error is a classical odd-error signal having both magnitude and sign or directional

information, blur is an unusual even-error signal having only magnitude and not sign or directional information. The importance of this in the behavior of the lens system will be discussed later.

3. The organized neurological networks in the retina and visual cortex then process the photoreceptor information so as to develop proper control signals to drive the final common path neurons in the Edinger-Westphal nucleus (37). Jampel (18) has shown increased refraction of primate eyes with occipital cortical stimulation. The ciliary ganglion relays the main parasympathetic fibers (35) although there is considerable debate regarding the presence or absence of orthosympathetic fibers and the presence or absence of adrenergic inhibitory synapses on the ciliary muscle. Further knowledge of the neurology of the lens system is in a primitive state.

4. The ciliary muscle was shown by Helmholtz (15) to contract during accommodation. This relaxes the tension applied via the suspensory ligament to the lens capsule. The capsule then elastically deforms and bulges the central anterior surface of the lens and this decreased radius of curvature increases the dioptric power of the eye. Clear vision position moves nearer. Conversely, when the ciliary muscle relaxes the suspensory ligament tenses and flattens the capsule, the lens decreases its power and the clear vision position moves

farther out toward infinity. Fincham (13) showed the equatorial diameter to be smaller in accommodation and Hess (16) showed that the loose suspensory ligament during accommodation permits gravity to affect lens position.

C. Older Physiological Behavioral Studies

1. The basic dynamics of the accommodative system is shown in Fig. 9 (33) which shows responses of the human lens to (a) repetitive step changes in target distance, (b) sinusoidal changes in target distance. Also shown is (c) the noise or random fluctuation in refractive power with a constant target distance. This noise fluctuation can also be noted in Fig. 9a.

"The initial reaction time to a 1 diopter step signal either positive or negative is approximately 425 milliseconds but the corresponding value for repetitive data was 240 milliseconds for positive accommodation and 200 milliseconds for negative accommodations" (8). See also Campbell and Westheimer 1960 (6). Further discussion of dynamical factors will be postponed to the next section on control systems studies.

2. Many studies have indicated that presbyopia, or the farsightedness of the aged, results from a reduction in elasticity of the lens and capsule or to a weakness or change in the ciliary muscle. This subject requires further

study but an excellent review exists by Alpern 1962, especially pages 206-217 (2). Of interest is the use of a bright light by presbyopes. This constricts the pupil, increasing depth of focus, and thus reducing blur of target image on the retina. A common complaint is "My eyes are all right, it is just that my arms are not long enough."

3. The perceptual aspect of the accommodative system is most intriguing. Ordinarily the convergence mechanism driven by binocular disparity is synkinetic with the lens system; also of importance in the natural situation is monocular parallax. Changes of size and intensity of targets ordinarily correlate with distance and are useful perceptual clues.

Usually targets in nature follow long courses during which time their heading and velocity can be estimated and used to predict changes in target distance. In the laboratory slight lateral movements due to misalignment of target motion with respect to the optical axis are commonly picked up by trained subjects. Any predictability in sequencing of target motions is quickly and unconsciously learned by subjects. However, all of the above can be controlled to produce the laboratory condition of restricted monocular viewing conditions. How does the lens system then behave? What does this tell us about the signal information handling mechanisms?

First of all, using two different experimental methods it was observed that 50% of initial movements in tracking step changes of target were erroneous, requiring secondary corrections to finally lock onto target distance with correct clear-vision-position and thus with zero steady distance error (32)(36).

These experiments make unlikely suggestions (7,12) that complex perceptual clues such as astigmatism, spherical aberration, and chromatic aberration can easily or are usually employed even in the restricted laboratory conditions by trained observers. In any case the crucial experiment of producing 50% initial errors with zero final DC errors shows the system to lack a basic odd-error signal mechanism.

D. Control Systems Experiments

In 1962 two independent but similar research studies were published: "A servoanalysis of the human accommodative mechanism" by John H. Carter (9) and "The dynamics of the human lens system" by L. Stark, Y. Takahashi and G. Zames (33). Both used direct recording infrared optometers.

Carter's study showed some very interesting phenomena. The open loop response to small steps had a gain dependent upon instructions to the subject. Predictive ability was demonstrated in that initial responses had longer latencies than responses to later repetitive target steps.

Unfortunately he was unable to derive an adequate and realistic transfer function because the empirical data defined a regenerative system. Robson with less complete data also displayed a frequency response diagram (4).

Stark, Takahashi and Zames were limited by the small amount of semiadequate data available to them. These had been obtained in experiments during July of 1959 while Stark was a guest at the laboratory of Dr. Fergus Campbell. However, these experiments had been designed to reveal the basic nonlinear characteristics of the system. Earlier in Fig. 9a the step responses show stable large amplitude responses and a fluctuation suggestively like an instability oscillation. In fact, Campbell, Robson and Westheimer in 1959 (5) had shown that the low amplitude fluctuations had a high frequency peak at 2 cps defined by power spectral studies. Further this peak disappeared with approximations to open loop operating conditions.

The experiments showed important gain changes with input amplitude and a describing function analysis proved to be quite successful. Figure 10a shows that first the system was divided into a no-memory nonlinear element followed by a linear frequency dependent part. The closed loop data were transformed analytically and the describing function $\frac{M}{E}$ (E) obtained is shown in Fig. 10b (33). This then enabled them to obtain the linear element frequency response $\frac{C}{M}$ (f)

plotted as the solid lines in Figure 11 of the next section. These data led to the construction of the model discussed in the next section.

E. Models

The transfer function of the linear part of the lens system was determined to be

$$F(s) = \frac{4}{s} \cdot \frac{(1 + 0.15s)e^{-0.1s}}{1 + 2[0.3(0.08s)] + (0.08s)} \quad (1)$$

where the complex pole pair with a natural frequency of 12.5 radians per second may be considered related to the mechanical system composed of the elastic capsule of the lens, suspensory ligaments and scleral restoring force and the elastic and viscous properties of the ciliary muscle. The gain curve of the transfer function (dashed heavy line) in Fig. 11 shows good agreement with experimental results (solid heavy line).

The nonlinear part represented by the describing function shown in Fig. 10b can be also represented by the input-output curve of Fig. 10c. This is unusual in that the output actual decreases as input increases over part of its domain. However, on consideration of the fact that the error signal is blur and that a more smeared image may well be a weaker error signal, one may find reasons for this

peculiarly shaped nonlinearity. Further there is an experimental function (14) showing human static accommodative responses for various DC inputs which resemble the curve of Fig. 10c.

The combined model accounts for large and small signal responses, stability characteristics and predicts certain noise spectral features of the open and closed loop lens system.

F. Critical Summary

Certain important discrepancies remain to be resolved by future experiments. The phase agreement of the model (light dashed lines) and experiment, (light solid line) in Fig. 11 is poor at low frequencies. This together with the phase lead element in the transfer function, the input adaptive prediction resulting in shorter time delays, and the lack of evidence of even-error signal behavior all suggest that prediction played a role unsuspected at the time in our 1959 experiments.

Much interesting development lies immediately ahead in the study of the accommodative mechanism. Better instrumentation, in particular a reliable optometer, is greatly needed. The interaction between accommodation and vergence movements will certainly throw a revealing light on the nature of strabismus, a most common visual ailment.

Even with our present limited knowledge, we might suggest an explanation for night or space myopia. It is known that instead of focussing at infinity, subjects have a certain amount, circa 0.4 diopters, of accommodation. This may be responsible for severe defect when jet pilots scan for enemy planes, for example. The noise in the accommodation system may produce this effect by subtracting its mean level value from infinity. These and other problems will be resolved as the conceptual methods of cybernetics are effectively applied with adequate instrumentation and with appropriate attention to physiological complexity.

The neurophysiological basis of the retinal and visual cortical operations are completely or almost completely unknown. Figure 12 from Polyak (24) shows the kind of highly structured and connected neural elements in the human retina. It is most important that decisive theoretical and experimental investigators in this "neural operator" area of research be stimulated by the unsolved questions above.

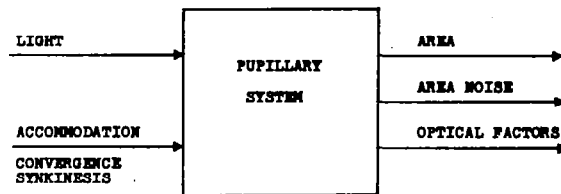


Fig. 1. Block diagram. Input-output identification for the pupillary system.

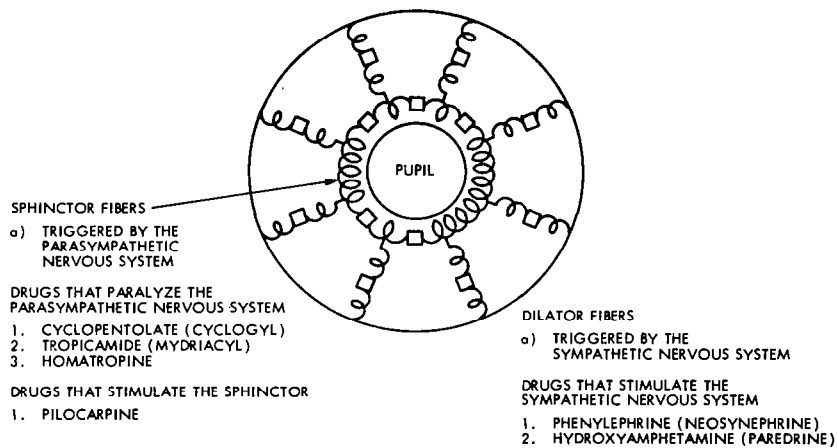


Fig. 2. A mechanical diagram representation of the sphincter and dilator muscles of the iris. (23)

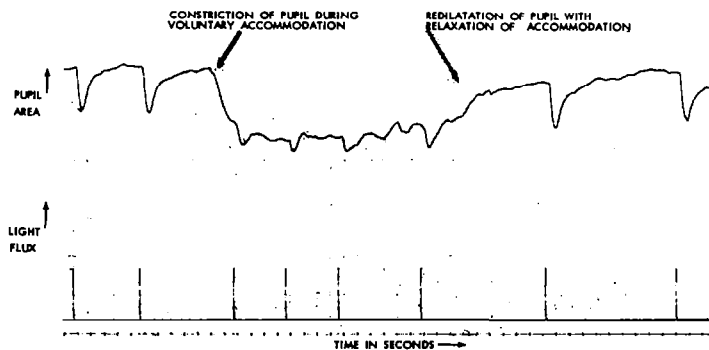


Fig. 3. Constriction and redilatation of the pupil during accommodation. (29)

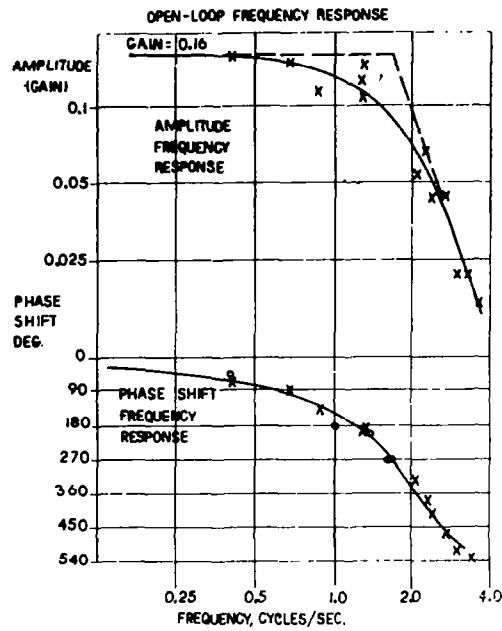


Fig. 4. Open loop frequency response. (30)

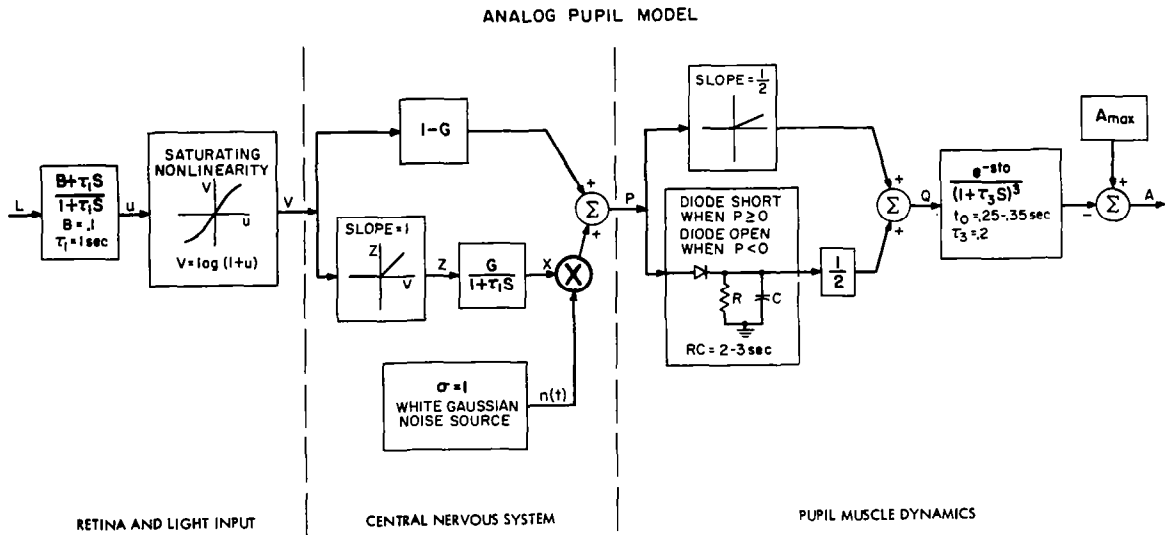


Fig. 5. Analog pupil model. (26)

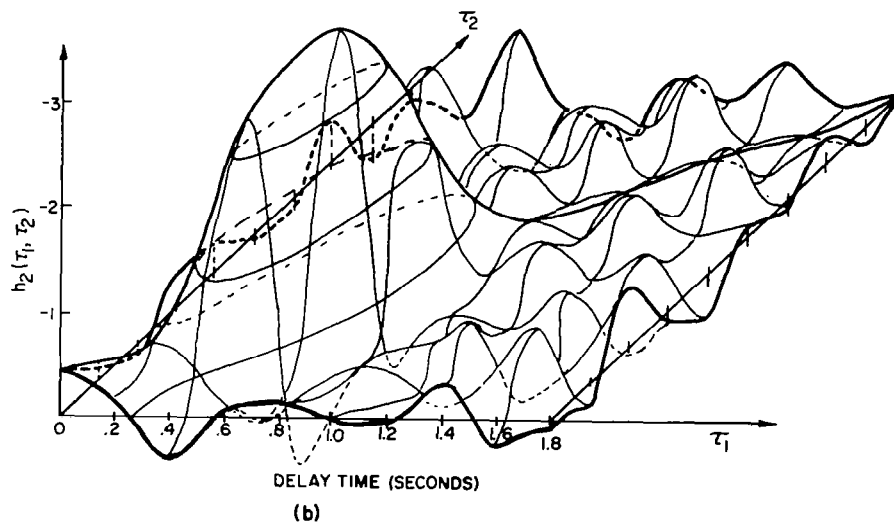
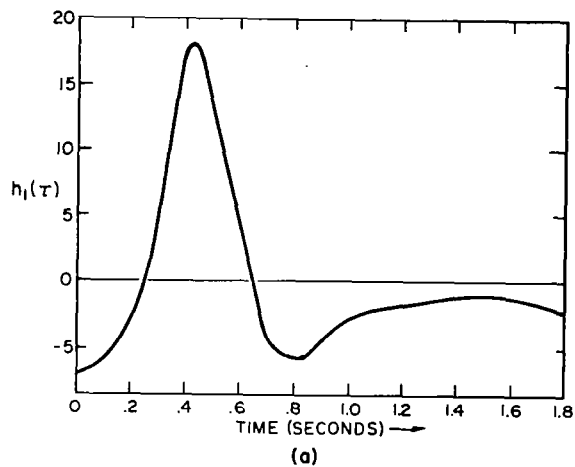
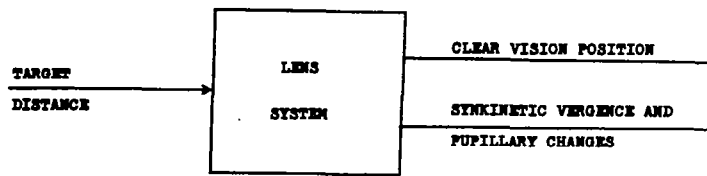
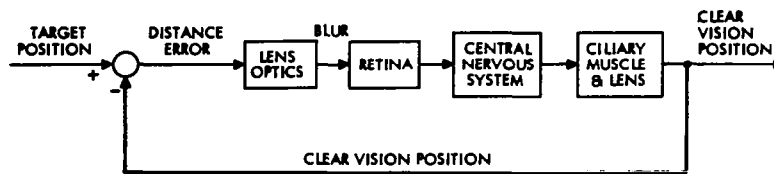


Fig. 6a & b. First and second order kernels of open loop pupil light servomechanism. (28)



(a)



(b)

Fig. 7. (a) Block diagram input-output identification for the lens system. (b) Block diagram of the human accommodation system. (33)

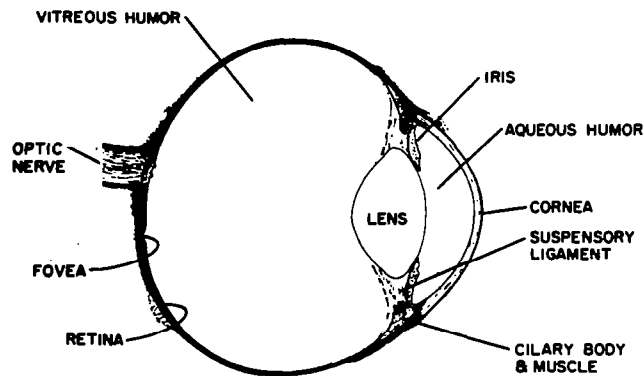


Fig. 8. Cross section of the human eye. (11)

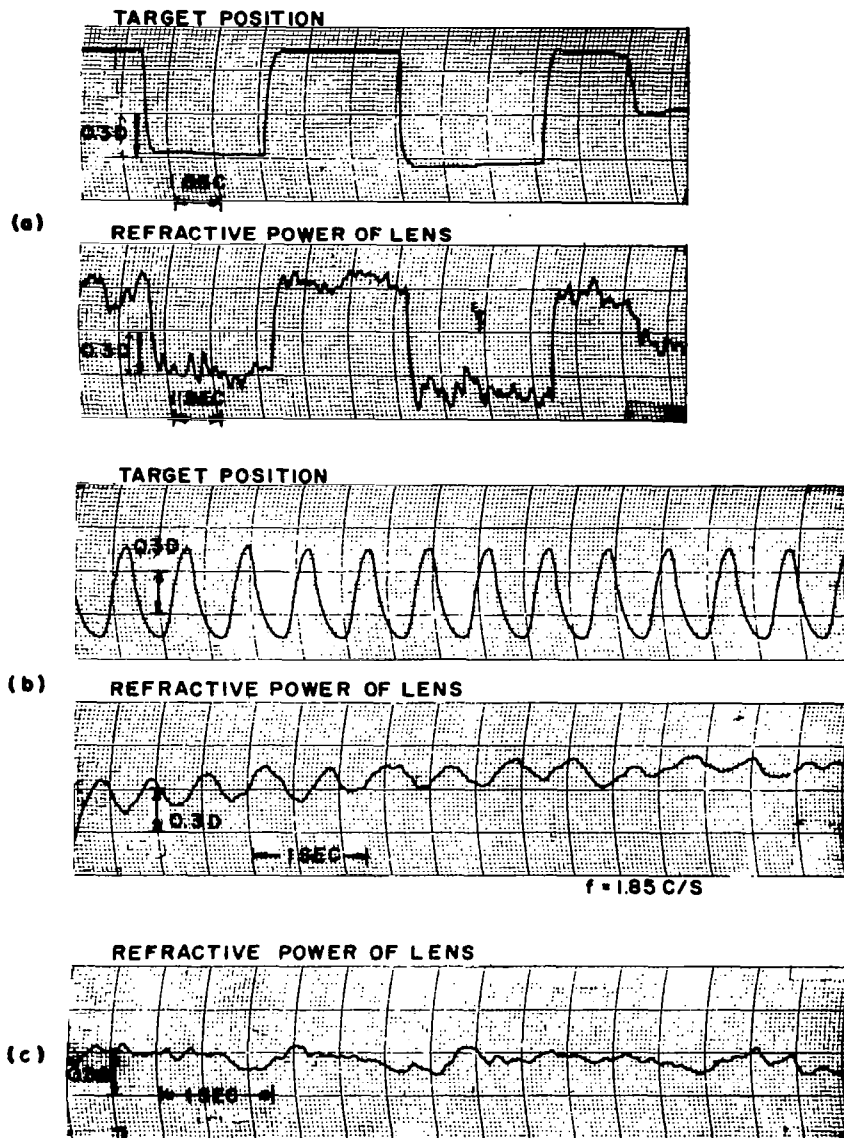


Fig. 9. Responses of the human lens to (a) repetitive step changes in target distance, (b) sinusoidal changes in target distance. (c) The noise or random fluctuation in refractive power with a constant target distance. (33)

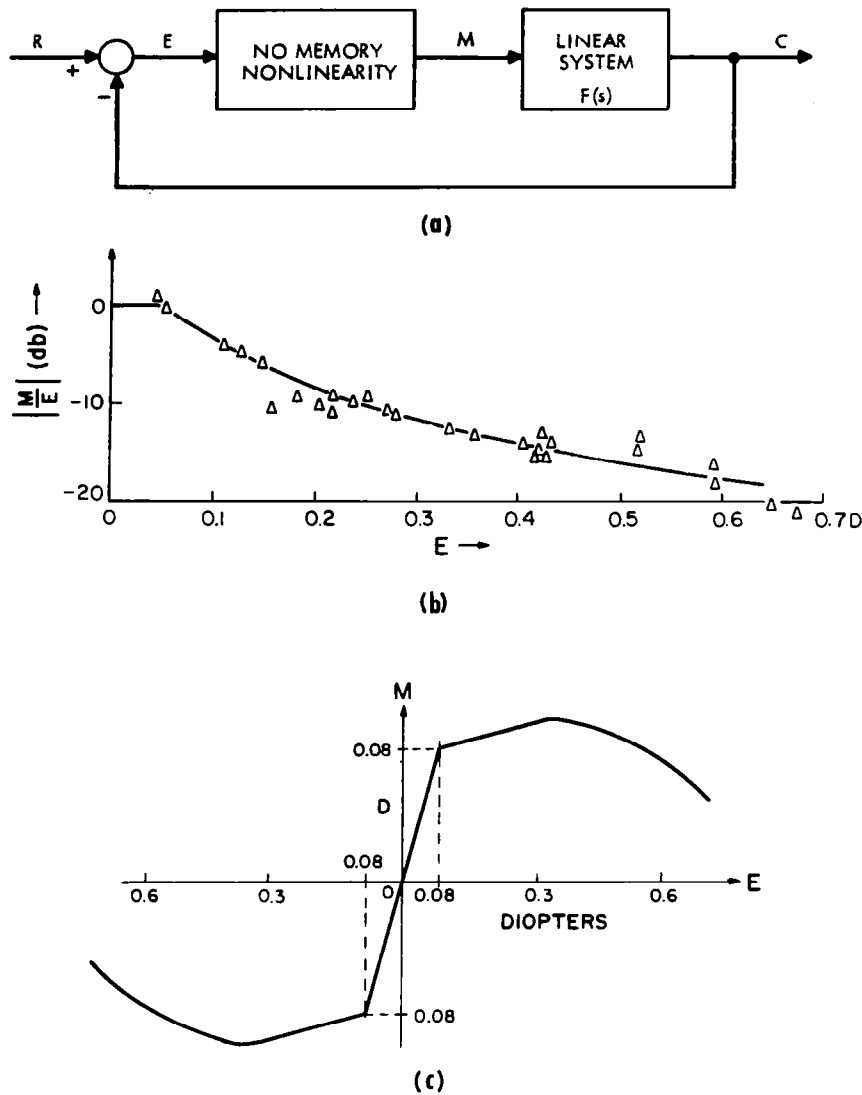


Fig. 10. (a) Lens control system divided into no-memory nonlinear and frequency-dependent linear elements. C is controlled quantity, clear vision position; R is reference input, target position; E is error or actuating signal, position error; M is signal representing output from nonlinear element and input to linear frequency-dependent element. (b) Describing function: $|M/E|$ as a function of $E \rightarrow$, obtained by separation of no-memory non-linearity from frequency-dependent linear function. (c) Model nonlinear relationship between input, e , and output, m ; small-signal proportional region, moderate signal saturation region, and even larger-signal region shows reduction of output. (33)

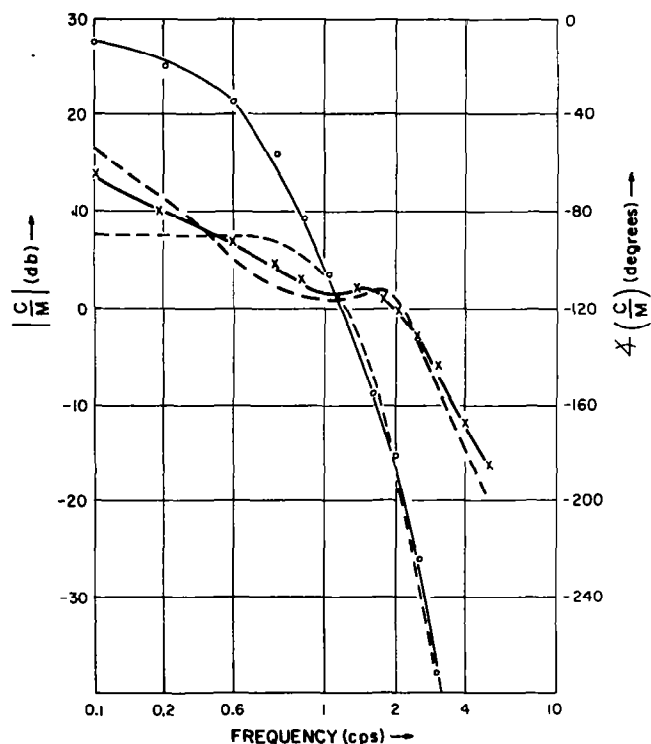


Fig. 11. Linear-element transfer function, $\frac{C(f)}{M(f)}$.
 Computed results from experiment are shown in continuous lines; heavy (crosses) from gain & thin (open circles) for phase. Model gain (thick interrupted line) & phase (thin interrupted line) from eq. (1) are also shown for comparison. Note good gain agreement; poor phase agreement at low frequencies suggesting a prediction operator, e^{+st} , effective in that range. (33)

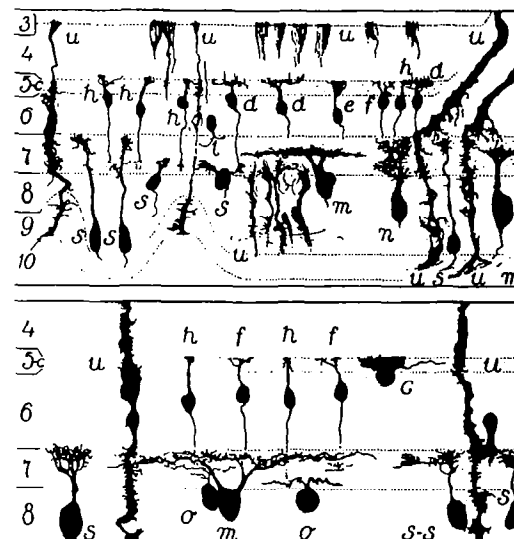


Fig. 113 Elements of human retina stained with the Golgi method. Upper sketch from a six-month-old fetus; lower sketch from one eight months old. Labeling: u, horizontal cell; d, map bipolar; e, brush bipolar; f, flat-top bipolar; h, modest bipolar; s, amacrine cell; m, parvocellular ganglion; n, small ganglion with horizontally running dendrites; s, modest ganglion; s-s, broadly distributed modest ganglion; s, radial fibers of Müller or their top ends, with so-called "horizontal baskets." At the right, in upper sketch, radial fibers have already assumed a slanting course, as found in the fully developed fovea and close to it. Note characteristic features by which each neuron variety may be recognized, even though the neurons have still far to go to assume their final form. The notable feature at the earlier age is the "diffuse character" of neurons. Here many collateral branches, thus in h and s, are still present, which later become discarded. In the advanced age the neurons almost attain their final, streamlined form. These gradual changes are to be interpreted as a functional specialization of various types of neurons in the process of maturing, which appears to be the essence of growth. (37, following figure.)

Fig. 12. The highly structure and connected neural elements in the human retina. (24)

References

1. Abelsdorff, G. Die Änderungen der Pupilleweite durch verschieden farbige Belichtung. A. Psychol. Sinnesorg. 22:81-95, 1900.
2. Alpern, M. "Accommodation" in The Eye, Vol. 3, ed. by H. Davson, Academic Press, N.Y. and London, pp.191-229, 1962.
3. Apter, J.T. Responses and distribution of the human intraocular muscles. AM.J. Ophthal. 49:1178-1192, 1960.
4. Campbell, F.W., J.G. Robson and G. Westheimer. Proc. of 2nd Congress International de Cybernetique, Namur, Sept 3-10, 1958, 1:924, 1958.
5. Campbell, F.W., J.G. Robson and G. Westheimer. J. Physiol. 145:579, 1959.
6. Campbell, F.W. and G. Westheimer. Dynamics of accommodation response to the human eye. J. Physiol. 151:285, 1960.
7. Campbell, F.W. and G. Westheimer. Factors influencing accommodation responses of the human eye. J. Opt. Soc. Am. 49:568, 1959.
8. Carter, J.H. An electro-chemical model of the human accommodation system. Optometric Weekly 54(16) 717-718, April 18, 1963.
9. Carter, J.H. A servoanalysis of the human accommodative mechanism. Arch. Soc. amer. oftal. optom. 4:137-168, 1962.
10. Clynes, M. Computer dynamic analysis of the pupil light reflex: a unidirectional rate sensitive sensor. 3rd International Conference on Med. Elec., London, 1960.
11. Cole, A., ed. Advances in Theoretical and Experimental Biophysics, Vol. 1, "Biological Control Mechanisms: Human Accommodation as an Example of a Neurological Servomechanism" by L. Stark, Fig. 1. Marcel Dekker, New York, 1965.
12. Fincham, E.F. The accommodation reflex and its stimulus. Brit. J. Ophthal. 35:381, 1951.
13. Fincham, E.F. The mechanism of accommodation. Brit. J. Ophthal. Suppl. 8:5-80, 1937.

14. Hamasahi, D., J. Ong and E. Marg. The amplitude of accommodation in presbyopia. Am. J. Opt. Arch. Am. Acad. Opt. Monograph 192, 1956.
15. Helmholtz, H. Uber die Accommodation des Auges. Graefes Arch. Ophthal. 1 (abt 2) 1-74, 1855.
16. Hess, C. Beobachtungen ueber den Akkommodationsvorang. Klin. Mbl. Augenheilk. 42:309-315, 1904.
17. Hubel, D.H. and T.N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. 160:106-154, 1962.
18. Jampel, R.S. Representation of a near-response on the cerebral cortex of the Macaque. Am. J. Ophth. 48 (No. 5, Pt. 2) 573-582, 1959.
19. Lettvin, J.Y., H.R. Maturana, W.S. McCulloch and W.H. Pitts. What the frog's eye tells the frog's brain. Proc. IRE Bio-Medical Electronics Issues, Nov. 1959, pp. 1940-1951.
20. Lowenstein, O. and I. Givner. Pupillary reflex to darkness. Arch. Ophthal. 30:603-609, 1943.
21. Lowenstein, O. and I.E. Lowenfeld. "The Pupil", in The Eye, Vol. 3, ed. by H. Davson, Academic Press, N.Y. and London, pp. 231-267, 1962.
22. Magoun, H.W. and S.W. Ranson. The afferent path of the light reflex. A review of the literature. Arch. Ophthal. 13:862-874, 1935.
23. Miller, D. and L. Stark. Effect of mydriatic drugs on pupillary dynamics. QPR 74, Res. Lab. Elec. MIT pp. 265-269, July 1964.
24. Polyak, S. The Vertebrate Visual System. Univ. of Chicago Press, p. 249, 1955.
25. Simpson, J.I. Kinematics and muscles of the human iris. M.S. Thesis, Dept. of Mechanical Engineering, MIT, June 1963.
26. Stanton, S. and L. Stark. Noise in the pupil. In preparation, 1965.
27. Stark, L. Biological rhythms, noise and asymmetry in the pupil-retinal control system. Ann. N.Y. Acad. Sci. 98:1096-1108, 1962.

28. Stark, L. Black box description and physical element identification in the pupil system. QPR 69, RLE MIT, p. 247, April 1963.
29. Stark, L. Nonlinear operator in the pupil system. QPR 72, RLE MIT, pp. 258-260, Jan. 1964.
30. Stark, L. Stability, oscillations and noise in the human pupil servomechanism. Proc. IRE 47:1925-1939, 1959.
31. Stark, L. and P.M. Sherman. A servoanalytic study of the consensual pupil reflex to light. J. Neurophysiol. 20:17-26, 1957.
32. Stark, L. and Y. Takahashi. Absence of an odd-error signal mechanism in human accommodation. Proc. IEEE International Convention, N.Y., pp. 202-214, March 1965.
33. Stark, L., Y. Takahashi and G. Zames. Nonlinear servoanalysis of human lens accommodation. Neurology Section, Electronic Systems Laboratory, MIT QPR 66, p. 337, 1962.
34. Stegeman, J. On the influence of sinusoidal variations in brightness upon pupillary size. Pflug. Arch. ges. Physiol. 264:113-122, 1957.
35. Stotter, W.A. The innervation of the intrinsic muscles of the eye; an experimental study. Proc. Soc. Exp. Biol., N.Y. 36:576-577, 1937.
36. Troelstra, A., B.L. Zuber, D. Miller and L. Stark. Accommodative tracking: A trial-and-error function. Vision Res. 4:585-594, 1964.
37. Warwick, R. The ocular parasympathetic nerve supply and its mesencephalic sources. J. Anat. London, 88:71-93, 1954.
38. Wiener, N. Nonlinear Problems in Random Theory. The Technology Press, MIT, and John Wiley and Sons, Inc. N.Y., 1958.
39. Young, F.A. An attempt to obtain pupillary conditioning with infrared photography. J. Exp. Psych. 48: 62-68, 1954.
40. Young, F.A. and W.R. Biersdorf. Pupillary contraction and dilatation in light and darkness. J. Comp. Physiol. Psychol. 47:264-268, 1954.
41. Young, T. On the mechanism of the eye. Phil. Trans. B. 23-88, 1801.

SECTION 4

CONTROL SYSTEMS AND BODY FLUID REGULATION

SUMMARY

The state of our knowledge regarding the biological control systems which operate to regulate the volume and composition of physiological fluids, has been heretofore largely empirical and intuitive.

A large body of empirical data has been generated by experiments designed to illuminate different single aspects of this regulation; the visualization of an overall design philosophy has, however, not been possible because of the large amount of experimental monitoring and computation involved.

It is possible, by applying certain concepts from control systems theory as used in engineering, to alleviate this situation. We have attempted to indicate in this report that much of the empirical data in fluid and electrolyte physiology can be conceptualized in a theoretical control systems framework.

Such a conceptualization of the system indicates the desirability of the reinterpretation of certain data, and perhaps even the revision of many basic concepts underlying previous formulations of fluid and electrolyte dynamics.

These considerations also suggest that simulation of the system, using mathematical programming techniques, is a realizable goal.

CONTROL SYSTEMS AND BODY FLUID REGULATION

by

Robert Taub, M.D.
and
Arthur Taub, M.D., Ph.D.

1. A REVIEW OF CERTAIN BASIC ASPECTS OF THE REGULATION OF BODY FLUIDS

BODY FLUIDS: COMPOSITION AND PARTITIONS

Introduction

A great deal of information regarding the structure and composition of the body fluids has been accumulated throughout the past 25 years. Most of this information has heretofore dealt with the cellular mechanisms and physico-chemical reactions responsible for the transport of one or another component of body fluids. In order, however, to visualize even a portion of the factors which regulate body fluid dynamics as a whole, it is necessary to develop a unified conception of body fluids, their individual interrelationships with one another then assuming a secondary importance.

Body Fluids: Definitions

The body fluids may be thought of as consisting of a solution of a number of solutes in a solvent, water. The

solution phase of the constituents of the organism is concerned with the transport of materials related to cellular metabolic activities. Transport occurs not only between cells, tissues and the internal environment, but is also performed within cells and between organelles. Thus, many of the solutes dissolved in water, as initial or end products of metabolism, are continually in motion between the internal environment and the organism, between the body fluids surrounding the cells and the interiors of the cells, and within the cells themselves.

It is sometimes helpful to think of the body fluids as providing the environment in which the transport function can take place. Although all of the components of body fluids are subject to turnover, some of the components have a slower turnover rate and are therefore considered to be relatively stable (e.g. the serum globulin, total protein, and others, under certain restricted conditions). (But it must be borne in mind that the concept of the body fluids as a relatively static structure or truly constant environment is not, in fact, valid.) Occasionally, it may be convenient to treat a portion of the solvent, water, as if it itself were a solute in solution, the same statement holding for water as for other constituents.

The turnover of all the constituents of body fluids provides, under appropriate circumstances, the opportunity for rapid change in both volume and composition and

concentration of these fluids, and also the ability to respond to physiological and pathological stimuli. In addition, a portion of the body fluids is in constant circulation (the intravascular portion); one of the characteristics of this system, is that a given portion of the body fluids, recirculated is iteratively manipulated . Thus an operation with little effect (such as the filtering of sodium through a glomerulus) may be amplified many fold. Although it is usually easy to detect experimentally the relatively large net effect thus produced, it is generally difficult to define the actual operation performed.

The maintenance of a constant composition and volume of the body fluids in the face of turnover of all components and the addition of metabolic and environmental loads, depends upon the constant addition of energy to the open system. Although a portion of the energy to drive this system is derived from the pumping action of the heart, energy is generally derived from the biochemical breakdown of energy-rich phosphate and the subsequent release of high energy bonds from compounds such as adenosine triphosphate. The sequence of mechanisms which the organism utilizes to supply this energy are complex in detail and are the proper province of biochemical energetics.

Body Fluid Phases and Compartments

The fluid phase of the body can be divided into two principal phases, the extracellular and intracellular. The extracellular phase is defined as that portion of the body fluids which lies outside cell membranes. The intracellular phase is that portion which lies within cells.

Extracellular fluid. Although by definition the portion of the body fluids that lies outside cell boundaries is considered to be extracellular fluid, the "functional" cell boundary may differ from the anatomical cell boundary. The volume of extracellular space is usually estimated by instilling into the intravascular space a substance which will assume a uniform distribution within all extracellular compartments and is assumed not to penetrate into other compartments. The volume of the extracellular compartment can then be calculated from the equation

$$V = \frac{A-E}{C}$$

where

A = the amount of material instilled into the volume,

E = amount lost during the time necessary for complete mixing,

C = concentration of the substance after complete mixing has occurred.

Many substances have been used in this manner to measure the volume of the extracellular space, including inulin, sucrose, mannitol, sulfate and chloride ion. Different values are yielded by these agents because a large proportion of dense connective tissue and bone is not readily penetrated by some of the larger molecules. Estimates for the total volume of the extracellular phase of the body vary from 16 to 25% of the total body weight. As a precise term, therefore, the "extracellular space" has little meaning except when used together with the solute used to measure its volume, and the technique involved.

Subdivisions of the extracellular space. The extracellular phase can be further subdivided into three principal compartments: the plasma, the interstitial fluid, and the connective tissue. The plasma is that portion of the extracellular fluid (the formed blood elements neglected) which is contained within blood vessels. The interstitial fluid is that portion of the extracellular fluid which is in contact both with the plasma and with the boundaries of cells. The connective tissue fluid constitutes that very minor fraction of the interstitial fluid which is in close contact with dense connective tissue and bone. Its composition varies little from that of interstitial fluid, except that its components probably have a lower turnover rate. The plasma and the interstitial fluid are separated by the walls of blood vessels, which constitute a semipermeable

membrane through which small molecular species can diffuse almost freely. For this reason plasma and interstitial fluid do not differ greatly in their electrolyte composition, nor in their osmolarity. Plasma, however, contains approximately 6 to 8 grams percent of protein of large molecular size which is not freely diffusible across the vascular wall. The protein content of plasma is therefore much higher than that of interstitial fluid. Because of the presence of a nondiffusible solute within the vascular wall the ionic concentration will differ slightly in plasma and interstitial fluid (as a result of the Gibbs-Donnan electrical neutrality constraint). Thus, anions other than multicharged protein molecules will be in lesser concentration in the plasma than in the interstitial fluid.

Intracellular fluid. Because of the high degree of physiological reactance associated with the investigation of the interior of individual cells, the composition of intracellular fluids cannot be studied directly and is therefore poorly understood. Certainly that proportion of water which is bound to the cytoplasm and that which is found within the nuclear sac bound to electrolytes, has not been determined. Furthermore, the water and electrolyte composition of cells belonging to different tissues will certainly differ, as a result of their functional difference. The intracellular water and electrolyte composition of tissues may differ from time to time because of the deposition and

mobilization of glycogen, fat and other metabolites in and from the cell. Since skeletal muscle represents the bulk of soft tissue in the body, various methods have been applied to determine a prototype composition of intracellular fluid. These methods, all of which are cumbersome, include that of performing determinations of mineral on fat free, dry solid tissues, and the use of indicators which freely permeate the membrane in their undissociated form but not in their ionized form (e.g. 5-5-dimethyl-2-4-oxazolidinedione). Using these techniques it has been inferred that the average composition of muscle cell fluid might be;

sodium - 10 meq/l
potassium - 150 meq/l
magnesium - 40 meq/l
bicarbonate - 10 meq/l
phosphate & sulfate - 150 meq/l
proteins - 40 meq/l.

Obviously an analysis of this type does not tell us the size, complexity, and number of molecular species in which phosphate, sulfate and proteins are incorporated, and yet these features are crucial with respect to information concerning the chemical reactions and equilibria in which they participate.

Practically, the principal facts to be kept in mind about intracellular fluid include the high potassium content, the relatively low sodium content, and the high content of protein, phosphate and sulphate. It is recognized that much of the potassium within the cell can be made to

leave the intracellular fluid under circumstances of severe potassium depletion; also hydrogen ions and sodium can and do penetrate cell membranes although their usual concentrations in intracellular fluids are quite low. The maintenance of the high concentration gradient for potassium and sodium depends upon an expenditure of energy ("the sodium pump") and therefore any stress on the system which disturbs this provision of energy can be expected to alter the relationship between the ionic content of intra- and extracellular fluids.

Water, on the other hand, is felt to be freely permeable through all cell membranes, and for this reason one would expect that water would distribute itself in such a manner that the total solute concentrations in each compartment of the body would be approximately equal, i.e., that there is "osmotic uniformity" between compartments.

Trans-cellular fluid. The trans-cellular fluid is that portion of the body fluids which is contained within cavities lined by epithelium and is thus not classified as either intra- or extracellular. These include the secretions of the digestive glands (pancreas and intestinal glands), the cerebro-spinal fluid, intraocular fluid and synovial fluid. The last three fluids are considered to be a transudate or an "ultrafiltrate" of plasma, which implies that the concentrations of small molecular species are virtually identical with that of plasma, but that protein is excluded.

The measurement of actual vs. theoretical volumes. As was indicated earlier, the volume of any particular compartment can be determined by allowing it to equilibrate with an indicator. In practice, this method is applicable only to plasma volume determinations with indicators such as T-1824 and radio-iodinated albumin (which are bound to proteins) and to the extracellular fluid volume measurements noted previously. It is not applicable to measurement of connective tissue water since this water has a slow turnover rate and mixing is therefore not complete until long after administration of the indicator. It is not applicable to measurement of intracellular fluid volume for obvious reasons. The volumes of interstitial fluid and intracellular fluid can be obtained by subtraction, however. Thus, the interstitial fluid volume would equal the total extracellular volume minus the plasma volume; whereas the intracellular fluid would equal the total body water (as measured by deuterium or tritium labeled water) minus the volume of extracellular fluid.

Constituents and Composition of the Various Phases of Body Fluids

As has been indicated, the composition of the fluid within the individual compartments reveals distinctive characteristics. The differences between plasma and interstitial fluid are for the most part due to passive forces,

i.e., energy which is not expended at the surface of the vascular membrane which separates the plasma water and the interstitial water. The electrolyte composition is similar except for the Gibbs-Donnan effect because of the decreased permeability of protein. The greatest difference in electrolyte composition is between interstitial and intracellular fluid. The predominant cation in interstitial fluid is sodium and in intracellular fluid is potassium. The predominant anions of the interstitial fluid are chloride and bicarbonate; of the intracellular fluid, phosphate and protein.

Principal cations. POTASSIUM. The functional integrity of the cell is dependent on an adequate potassium intake; no other ion can functionally replace potassium in the cell. Potassium is readily absorbed from the intestinal tract and absorption is nearly complete. Certain fluids such as gastric and intestinal secretions contain a higher concentration of potassium than does the extracellular fluid. Normally around 100 meq. of potassium are ingested in the daily diet; most of it is excreted in the urine. In response to the rise of concentration of body potassium renal excretion of this ion is increased. During this process, the concentration of potassium in the extracellular space falls slightly, and potassium shifts from an intracellular to an extracellular site. Although it is probably true that the amount of potassium which enters cells is due to an active transport

mechanism which serves to transport potassium against a concentration gradient, nevertheless the amount which does enter the cell depends to a certain extent on the concentration of potassium in the extracellular fluid; this concentration must be kept within strict limits (2-5 meq/l) to obviate effects on cardiac electrochemistry which may be deleterious.

SODIUM. Sodium is the chief cation of the extracellular fluid. The average intake of sodium chloride is approximately 10 grams/day; approximately this amount is excreted in the urine. Osmolarity is thus maintained. There are other minor pathways of excretion which include the skin and the gastrointestinal tract. The total amount of sodium which is contained in the extracellular fluids is the principal determinant, under normal conditions, of the total volume of extracellular fluid. If large amounts of isotonic sodium are absorbed and cannot be excreted, the plasma and interstitial fluid volume expand accordingly to isosmotically accommodate the sodium. The conservation of sodium salt is a function which is performed primarily by the kidney. This mechanism involves the active transport of sodium and its reabsorption from tubular urine into the plasma after filtration.

CALCIUM. Calcium is present in the extracellular fluid to a concentration of between 9 and 11 milligrams/liter. Its presence in body fluids is necessary because it subserves a variety of functions relating to the integrity of bone and

of the cell membranes. Calcium is absorbed by active transport in the upper part of the small intestine; this active transport being under the partial control of vitamin D and vitamin D analogues. Parathyroid hormone also will increase intestinal absorption of calcium. Calcium circulates in the plasma in three forms: (1) bound to protein, (2) complexed with certain anions such as citrate, and (3) freely ionized. Ionized calcium is freely diffusible and comprises about 65% of the total blood calcium. The dynamics of renal excretion of calcium, and the factors controlling this excretion are not known at this time. An increased load of calcium to the kidney will result in increased renal excretion of calcium.

MAGNESIUM. Magnesium, like calcium, is an ion of considerable physiological importance. It is also essential for the functional integrity of the neuromuscular system. Magnesium is absorbed with difficulty from the gastrointestinal tract and circulates in the serum about 80% ionized. Magnesium is excreted mainly in the urine although a small amount may be eliminated into the intestinal tract. The kidney excretes magnesium so rapidly that the oral ingestion of magnesium salts results in only a slight rise in plasma magnesium. Magnesium concentrations are subject to factors which probably are not involved with the maintenance of the volume or composition of other constituents of the body fluids.

Principal anions. BICARBONATE. The bicarbonate ion is derived in the body from a reaction which involves the combination of carbon dioxide and water to form carbonic acid, which then dissociates to the bicarbonate ion and hydrogen ion.

PHOSPHATE. The phosphate ion serves a variety of functions in the body. These are concerned with the structure of bone, the acid base equilibrium of body fluids, the regulation of calcium metabolism, and a variety of intracellular functions relating to the production of energy. Phosphate is slowly absorbed from the gastrointestinal tract, circulates in the blood as dibasic and monobasic ions, (their relative proportions being determined by the pH of the blood) and is excreted by the kidney. At the normal pH approximately 80% of plasma phosphate exists as dibasic phosphate. The dibasic and monobasic phosphate ions serve as a very efficient buffer pair, but in the plasma the concentration of phosphate is not sufficiently high for this to be a buffer of great physiological importance. It is of some importance as a buffer in the urine. (This will be discussed in the section on renal physiology.) The renal excretion of phosphate is under the influence of both vitamin D and parathyroid hormone.

CHLORIDE. Chloride is the chief anion of the extracellular fluid. One of its major functions is to preserve osmotic and electrical equilibrium. There is evidence,

however, that the level of serum chloride is related inversely to the level of serum bicarbonate under pathological conditions. Chloride ion is rapidly absorbed by the gastrointestinal tract and is excreted primarily by the kidneys but also into the sweat and into the gastrointestinal tract as well.

EXCHANGE OF WATER AND ELECTROLYTES
BETWEEN VARIOUS COMPARTMENTS OF THE BODY

Types of Movements of Fluid Components of the Body

The components of the body fluids, both solvent and solutes, move along energy gradients that arise from the specific activities or chemical potentials of the components. There are four basic kinds of movement or transfers of water and electrolytes in the organism. These are: diffusion, osmosis, "active" transport, and mass movement due to hydrostatic pressure.

Diffusion. Diffusion along concentration gradients is the simplest type of fluid movement and is usually produced by a change in concentration as a result of the addition or subtraction of solute or solvent. The net result of shifts of solute and solvent due to these processes will depend upon the magnitude of diffusion of the solutes and solvents involved. The equilibrium that is finally reached, however, cannot always be predicted by simply considering the concentration gradient, without considering the diffusion rates of the solute and the solvent. For example, if one introduces a hypertonic solution of glucose into the interstitial fluid the end result of such an infusion will be the complete absorption of both the water and the glucose.

Although the capillary membrane is permeable

to both there will nevertheless first be a large outward shift of water from the blood vessels. This is due to the fact that the glucose will traverse the capillary membrane at a slower rate than the water; thus for a certain time period the solution outside the vessels will be hypertonic to plasma and a concentration gradient for water will be set up which will be nullified only when the glucose has been sufficiently absorbed. The energy operating this system is derived from the process initially producing the concentration difference.

Osmosis. Osmosis is the term used to indicate a concentration gradient created by the restraint of a solute present in solution to one of two fluid phases, usually separated by a semipermeable membrane. When these phases are differentiated not only by inequalities in concentrations of solutes and solvents, but also by the restraint of non-transferability, or low permeability of one solute in one of the phases, then the solvent will transfer between the two phases until their respective concentrations are equalized. The force exerted by this transfer of solvents is termed the effective osmotic pressure and may be expressed in equivalent units of hydrostatic pressure which are required to oppose it in a state of equilibrium. The energy which produces the osmotic pressure is derived from the process that led to the change in concentration of the restrained solutes. In this context, it must be mentioned that a change in overall permeability of one solute may not

only be due to the difficulty of diffusion across a membrane, but may also be due to active processes at the membrane site which require energy expenditure at that site and which will effectively actively transport the ion from one side of the membrane to another. Such a system probably exists at cell boundaries where sodium is actively transported out of the cell and in such circumstances a high concentration gradient exists for sodium across the cell membrane.

"Active" transport. Active transport is the movement of a solute against a concentration gradient. Such a transfer is called "active" because it requires expenditure of a further increment of energy at the site of transport. The energy in this system is usually derived from the breakdown of high energy phosphate bonds, as was mentioned earlier. The results of active transport of one solute upon the other solutes and solvents in the fluid phases separated by the actively transporting membrane will depend upon their diffusibility. If other solutes are freely diffusible they will simply diffuse in a direction opposite to that of the actively transported solute; if they are not, an osmotic gradient is then established with the resultant shift of solvents.

Hydrostatic mass movement of fluid. Mass movements of fluids due to hydrostatic pressure may occur even when there are no differences in concentration or transferability of solute or solvent, if the hydrostatic pressure on one side of

a semipermeable membrane is higher than on the other and therefore the specific activity or the chemical potential of all the constituents on that side of the membrane will facilitate transfer to the other phase along the pressure gradient, until the pressures are equalized. The energy in this instance is derived from the process that sets up the hydrostatic pressure gradient.

It must be borne in mind that any freely diffusible solvent or solute in the biological organism is in constant and rapid turnover. The flux or rate of movement in one direction of an individual ion or molecule between two phases is usually occurring simultaneously with its movement in the opposite direction. Any accumulation of a molecular or ionic species on one side of the membrane as the free energy of the system approaches a minimum is due to the amount by which the flux in one direction exceeds the flux in the opposing direction. The turnover of the ion is usually considered to be the absolute sum of the fluxes while the net flux is considered to be the difference between the fluxes in both directions. At equilibrium, although the rate of turnover may be high, the net transfer fluid or other components may still be zero.

Access of Water from the External Environment into the Extracellular Fluid

Water usually enters the organism via the gastrointestinal tract. Ingested water, being hypotonic to plasma, is rapidly absorbed across the lining of the gastrointestinal tract (a semipermeable membrane) by virtue of the concentration gradient of the hypotonic water. If hypertonic solutions such as sea water are ingested no absorption occurs and in fact there is a net flux of water from the extracellular fluid into the lumen of the bowel causing a net loss of water from the organism.

Water may also be given in the form of isotonic solutions by the intravenous route. This route of administration of fluids is important experimentally since it provides, in contrast to orally ingested water, the addition of a known amount of solute and solvent to a given compartment, the plasma.

Another highly unusual portal of entry of water into the organism is via the lungs, as in cases of drowning. These same osmotic considerations apply to fluid which is instilled into the tracheal and bronchial tree; where a hypotonic solution is instilled, absorption of water is extremely rapid and may be sufficient to cause a rapid death of the animal. Conversely, hypertonic solutions cause a rapid loss of water from the organism.

Thirst. Although ingestion of water may be cursorily regarded as an indiscriminate process, the quantity of water which is taken in by an organism is quite precisely regulated. Although the sensation of thirst is subjective, and therefore difficult to quantitate precisely, it is definitely related to certain other parameters which are essential for the regulation of body fluids. The relationship of thirst to body fluid regulation can be summarized as follows: Thirst may be regarded as a servomechanism having as its output the elaboration of a desire to drink water. The input to this system is contraction of the intracellular volume of certain cells (sensors) which are located in the hypothalamus. The contraction of the intracellular volume of these cells follows an increase in the concentration of nondiffusible solutes in the extracellular fluid. When the effective osmolarity of the extracellular fluid is thus increased (either because of an increase in effective solute concentration or water loss in the body) the sensation of thirst appears and the organism will drink water, thus lowering the effective osmolarity of the extracellular fluid, thus also completing the feedback loop.

Although cellular dehydration appears to be the primary stimulus to thirst there are other stimuli to thirst as well; these are thought to be related to contraction of the total extracellular fluid volume. This, however, is a second order effect.

Fluid Transfer between Extracellular Compartments

Since water and other solutes are freely diffusible across the vascular membrane (except for large molecules such as protein) in order for these materials to be transported from one compartment to another the addition of free energy is required. This energy is supplied as hydrostatic pressure, derived from the pumping action of the heart and from muscular activities in muscles surrounding smaller veins and lymphatics. Most of the exchange of small molecules between the plasma and interstitial fluids occurs through the capillary wall which is the thinnest part of the vascular system and presents the least barrier to diffusion. The region where rapid diffusion occurs extends from the arterial to the venous end of the capillary. In every area along this region some water molecules are continually moving from the inside to the outside of the capillary and vice versa. Since the capillary presents a definite resistance to the passage of water through it, it follows that there must be a pressure gradient, established by the pumping force of the heart from the arterial to the venous end of the capillary. Under normal conditions the pressure at the midpoint of the ordinary capillary is most tissues (except the lungs) is about 25mm of mercury. It is somewhat higher than this in the arterial end and somewhat lower in the venous end of the capillary. The hydrostatic pressure of the tissue fluid just outside the capillary varies considerably, but is normally only a few

millimeters of mercury. The net hydrostatic pressure would thus tend to produce a net flux of water out of the capillaries into the interstitial fluid compartment. At normal equilibrium this does not usually occur because of the osmotic pressure of the large protein molecules which are present within the vascular system. The net result of these three separate forces (hydrostatic pressure in the vessels, tissue pressure exerted upon the interstitial fluid, and colloid osmotic pressure) would tend to create the following situation: At the arterial end of the capillary the hydrostatic pressure exceeds the colloid osmotic pressure plus the hydrostatic pressure of the tissue fluid and transudation of fluid from the capillary into the interstitial fluid occurs. At the venous end of the capillary the hydrostatic pressure has been sufficiently reduced so that fluid is now reabsorbed from the interstitial space. This arrangement produces a circulation of fluid in the interstitial space surrounding each capillary.

The "colloid osmotic pressure" which results in an unequal chemical potential of water between the plasma and the interstitial fluid is the result of the presence of large molecular weight protein (at least 70,000 or more) within the blood vessels. Protein molecules of this size are so large that they are generally unable to pass through the pores of the capillary wall; at the same time the particle size is small enough so that a significant colligative effect is present. A small amount of protein actually does leak

through the capillary wall but the bulk of the plasma proteins are held back. The albumin molecule, which contributes slightly over half of the total weight of plasma protein provides many more colloid osmotic particles than any other class of plasma protein. This colloid osmotic pressure due to the presence of albumin and other proteins amounts to about 25mm of mercury.

There is usually a low concentration of protein molecules in the extravascular interstitial fluid which is due to leakage. These protein molecules in the interstitial fluid are continually absorbed by the lymphatics and are thus eventually returned to the blood via the thoracic duct. Protein molecules are able to pass more easily through the membranes of the lymphatics, and once they are present in these vessels are transferred through the lymphatic channels by very slight hydrostatic pressure gradients set up by contraction of the body's musculature adjacent to the lymphatic vessels. That movement of fluid proceeds in one direction is insured by the presence of numerous valves in the lymphatic channels which prevent backward movement.

Because of the ready access of water and proteins into the lymphatics it is to be expected that lymph differs very little in its composition from that of interstitial fluid. A number of studies have demonstrated that the total extracellular fluid may practically be considered as a single entity,

as far as water, small lipid-insoluble ions, and other molecules are concerned.

Although the average capillary pressure in most tissues of the body is about 25mm of mercury, this situation does not always obtain. In the lung capillaries the hydrostatic pressure is only about 9mm of mercury. Hence in the pulmonary capillaries the colloid osmotic pressure is much greater than capillary hydrostatic pressure. From these considerations it would appear that the minutest film of water surrounding these capillaries would be sufficient to maintain equilibrium with the plasma. This would also explain the great rapidity with which water poured down the trachea is rapidly absorbed such as in drowning. In the capillaries of the kidney, particularly the capillaries in the kidney glomerulus, the hydrostatic pressure is about 75mm of mercury and greatly exceeds the colloid osmotic pressure. Thus there takes place a continuous transudation of fluid from the glomerular capillary into the urinary tubules and this is the mechanism by which filtration of the blood is accomplished.

Fluid Exchange between Intra- and Extracellular Compartments

Because of the specialized nature of some tissues, particularly the kidney, the mechanisms effecting transfer of fluid between intracellular and extracellular compartments can be expected to differ for different tissues. In general,

the exchange of electrolytes across the membranes separating intracellular and extracellular fluid is poorly understood. It is known that water is freely diffusible across the lining membranes of cells. Probably there is no concentration gradient for water between the intra- and extracellular fluid. Sodium and potassium are also quite permeable to the cell membrane, however, the movement of sodium and potassium ions between extra- and intracellular fluids is usually against a concentration gradient. The process by which these cells accumulate these solutes by energy requiring reactions is called active transport. There are several ways in which an ion may move from one side of the cell membrane to the other; diffusion may be along chemical or electrical gradients. Since an ion may move passively down an electrical gradient although the movement is against the chemical gradient; it is clear that a decision as to whether the net change is passive or active cannot be made without considering all the forces involved. The electrochemical potentials of two ions distributed passively on two sides of a membrane must be equal, and for a given ion this should approximately reduce to the Nernst equation

$$E_m = \frac{RT}{ZF} \ln \frac{C_i}{C_o}$$

where E_m is the electric potential difference across the membrane, R , T , and F are constants, Z is the valence of the ion, and C_i and C_o refer to the concentrations of

the ion inside and outside the membrane. One usually therefore infers that active transport is taking place when the membrane potential does not agree with the concentrations of this ion that one would expect on both sides of the membrane. Since the cell membrane is completely permeable to water and the cell contains large ions such as proteins which cannot leave the cell, the volume of the cell is critically dependent on the mechanism responsible for this steady state distribution of ions by active transport. Thus the active transport of these solutes regulates the life of the cell, and in this sense may be said to regulate the water exchange both inside and outside of the cell.

Bone

It might be helpful to include some remarks about the relationship of sodium and potassium which are deposited in bone to the maintenance of the ionic composition of body fluid. The composition of bone differs strikingly from that of other tissues in that only about one-fifth of bone is water and in that only about one-third of bone solids is protein. Two-thirds of the solid components of bone are composed of inorganic salts which are organized in a crystal lattice structure. Approximately one-third to a little less than one-half of the total body sodium is present in bone; of this

only about 15% can be ascribed to the extracellular fluid and less than 1% to the intracellular phase of bone. The major fraction is believed to be part of the crystal structure of the bone cells, and approximately 30 to 40% of this is exchangeable with an injected circulating isotope of sodium within 24 hours of its administration. There is considerably less potassium than sodium in the mineral phase of bone. It has been demonstrated that sodium and perhaps potassium deficits are shared by the bones, and that these cations appear to exchange for hydrogen ions in opposing distortions in acid base relationships. This area is still not well studied; it may be that in future studies of electrolyte composition the presence of a large reservoir of sodium which is slowly exchangeable with the remainder of the total body sodium may have to be taken into account.

The Extrarenal Excretion of Water and Electrolytes

There are several pathways by which water and electrolytes may be lost from the organism into the external environment. These include "insensible" perspiration, "sensible" perspiration, and gastrointestinal tract secretions.

Insensible perspiration. This term refers to the evaporative loss of water without loss of solute, from the expired air in the lungs and from the surface of the skin. Inspired air is exposed to a large surface in the pulmonary alveoli where its temperature is raised to that of blood and

it becomes almost saturated with water vapor. This rate of water loss in the lungs is relatively constant and is under no known modifying or controlling influences. The rate of water loss from the skin is conditioned by the surface area of the skin, the difference in temperature between the evaporating surface of the skin and the environmental air, and the vapor pressure of the environmental air. There is some evidence also that the rate of insensible water loss from the skin is inversely correlated with the concentration of sodium in the extracellular fluid. Since water is lost from the skin in this fashion as vapor, it would follow that the loss would be regulated in part by the chemical potential of the water of the body's fluids.

The total insensible loss per day is between 600 and 1000 millilitres in an average adult. This amount of water loss also accounts for the loss of about 25% of the heat produced in the body each day. The interrelationships between insensible loss of water and heat are further emphasized when one notes that the rate of insensible water loss is accelerated by an increase in the metabolic rate, as in exercise or fever.

Sensible perspiration. This differs from insensible perspiration in that solutes are excreted as well as water. Sweat is always hypotonic in healthy individuals. The average

composition of sweat is as follows:

sodium	48	millimoles/l.
potassium	5.9	millimoles/l.
chloride	40.0	millimoles/l.
ammonia	3.5	millimoles/l.
urea	8.6	millimoles/l.

Several factors modify the production of sweat. The primary stimulus appears to be heat. Heat receptors are not located in the skin but rather internally, most likely in the anterolateral portion of the hypothalamus which regulates the motor activities necessary to eliminate body heat. Other factors which will modify the volume and composition of sweat include the effective osmolarity of the body fluids and the secretions of the adrenal cortex. Salt depletion induces a diminished concentration of sodium and chloride in sweat, the effect probably being mediated by means of adrenal cortical hormones.

Gastrointestinal tract. All of the gastrointestinal secretions are approximately isotonic with plasma, except for saliva which is hypotonic. Aside from the high hydrogen ion concentration in gastric secretions, the major cation of the intestinal secretions is sodium. The concentration of sodium, however, diminishes and reciprocates with potassium as the caudal end of the bowel is approached. Chloride is the major anion of the fluids in most of the upper tract secretions and bicarbonate is the secondary ion.

The total fluid volume secreted into the gastrointestinal tract is estimated to average 8 litres per day; however, since there is a more or less continuous process of secretion and reabsorption there is very little net exchange. The gastrointestinal tract contains 1.5% of the total body water, 1.6% of total exchangeable sodium, 1% of total exchangeable potassium and almost 2% of total exchangeable chloride. Because of the relatively large volume of secretion, however, accidental loss of these fluids may result in severe losses to the body, because of their constant rapid turnover.

The net exchange of water between the intestinal tract and the extracellular fluid is conditioned by the difference in chemical potential of the water in these two fluids; most probably the movement of water across the lining membrane of the intestinal tract is entirely passive. Specific ions, however, can be absorbed and excreted against concentration gradients so that it is likely that active transport or ion exchange mechanisms are involved in these processes. The transport of ions in the small and large intestine appears to be modified by hormonal influences, specifically the secretions of the adrenal cortex. The secretion of hydrochloric acid by the stomach depends both on hormones and on nervous secretions and requires in addition an enzyme (carbonic anhydrase), a source of energy, and certain metabolic substrates. Certainly there is a good deal of active chloride transport by the gastric mucosa.

MAINTENANCE OF ACID BASE EQUILIBRIUM

Introduction

The dynamic steady state of the body fluids is maintained with considerable precision with respect to its total concentration of solute, the volumes of each compartment, and the concentrations of individual molecular species, for a given individual. The constancy of this environment must be maintained to allow metabolic processes to operate in an optimum environment. This is certainly true with respect to the hydrogen ion concentration of pH of the extracellular fluid.

Paradoxically, an understanding of the factors which regulate the pH of body fluids is perhaps easier to achieve for those who are not clinicians since much of the clinical literature has been confused until recently by many different non-chemical approaches which served only to complicate the issue.

Buffer Systems

Physiological solutions are characterized, in part, by the presence of a number of poorly dissociated acids together with their more dissociated salts. The dissociation of these acids into their more soluble forms follows the Henderson-

Hasselbalch equation

$$\text{pH} = \text{pK} + \log(\text{BA}/\text{HA})$$

where

BA = the concentration of the salt,

HA = concentration of the acid,

pK = the logarithm of the dissociation constant
of the acid.

Biologic fluids contain several such buffer systems; these include poorly dissociable proteins, a buffer pair of the monobasic and dibasic phosphate ions, and the carbonic acid bicarbonate buffer pair. The ratio of the components of each of these buffer pairs is determined by the pH, and in turn the several relationships of this salt to its respective acid will determine pH. If the ratio of any one of several buffer pairs is altered, all of the ratios of the other buffer pairs must change ("isohydric principle").

The fluid which is most accessible for analysis is the plasma, and determinations of pH on the plasma approximate the pH of the extracellular fluid. pH may be estimated directly or may be computed from a knowledge of the relationship between, for example, the bicarbonate concentration and the sum of the dissolved CO_2 and carbonic acid in the plasma. This relationship is then fitted into the Henderson-Hasselbalch

equation as follows:

$$\text{pH} = \text{pK}'_{(\text{bicarbonate})} + \log \frac{\text{HCO}_3^-}{(\text{H}_2\text{CO}_3 + \text{Dissolved CO}_2)}$$

The denominator of this buffer ratio is written as such because the ratio of dissolved carbon dioxide to carbonic acid is approximately 800:1, and the quantity of dissolved carbon dioxide is proportional to the tension or partial pressure of carbon dioxide in the plasma. Since dissolved CO_2 is, in turn, in equilibrium with carbonic acid, the sum of the dissolved carbon dioxide and carbonic acid is proportional to the pCO_2 , and depending upon the pH, will be proportional as well to the concentration of bicarbonate.

The pK' for the bicarbonate- CO_2 system has been empirically determined to be 6.10 at body temperature. By neglecting the quantity of carbonic acid as compared with dissolved carbon dioxide, and by considering the quantity of dissolved carbon dioxide at body temperature to be equal to 0.0301 times the partial pressure of carbon dioxide (in millimoles per litre per millimeter of mercury) we then arrive at the equation

$$\text{pH} = 6.10 + \log \left(\frac{\text{bicarbonate concentration}}{0.0301 \text{ (partial pressure of CO}_2\text{)}} \right).$$

The bicarbonate concentration is not measured directly but is estimated as the difference between the total CO_2 content of

the plasma and the quantity of dissolved CO_2 , so that the final form of the equation as it is used to measure pH (or alternatively the concentrations of any of the other components) would be

$$\text{pH} = 6.10 + \log \left[\frac{\text{Total } \text{CO}_2 - 0.0301 \text{ pCO}_2}{0.0301 \text{ pCO}_2} \right]$$

At first glance it might be assumed that the bicarbonate carbonic acid buffer pair is highly inefficient at the pH of extracellular fluid since the ratio of the numerator and denominator is not close to one. However, this is not the case; the buffer is highly efficient, primarily because the pCO_2 and therefore the amount of carbon dioxide which is dissolved in the plasma can be altered drastically. The gas is ubiquitous since it is constantly being formed as a product of metabolism, and because it is a gas, its rate of elimination can be substantially altered by changes in the depth and rate of respiration.

It must be emphasized that the bicarbonate CO_2 buffer system is the easiest in the body to study but this does not imply that it is the only principal buffer in the body. The hemoglobin of the erythrocytes, the proteins of the plasma, and the phosphates, bicarbonates, and proteins of the intracellular fluids each make significant contributions to overall buffering activity. The sum of these buffers has

been referred to as "buffer bases" and its range of normal values is 46 to 52 millimoles per litre. For purposes of computation some information about the activity of these bases may be obtained by examining the interrelationships between concentrations of sodium and the sums of the concentrations of total CO_2 and chloride. Under ordinary circumstances the difference between the concentrations of sodium and the sum of the concentrations of the two major anions is between 5 and 10 millimoles per litre. Since the retention, or excessive elimination of carbon dioxide will be accompanied by a reciprocal change in the concentration of chloride, the difference continues to be approximately 5 to 10 millimoles per litre. In contrast, when there is a metabolic disturbance of acid base equilibrium and bicarbonate is displaced by some undetermined anion, the difference will be increased. The exact concentration of other anions which go to replace bicarbonate may include phosphates, lactates, and other ions which are not routinely measured clinically. Thus, when the pH is made to deviate significantly from normal, such as in acidosis or alkyllosis, the actions of the buffers will tend to minimize this deviation. In certain compensated states then, there may be marked changes in patterns of renal excretion of electrolyte and pulmonary excretion of carbon dioxide at times where the pH differs but slightly from normal.

Correlates of Extracellular Buffer Activity

It is to be expected that changes in pH may also be reflected by some form of buffer action in the intracellular fluids. Because of the difficulty in diffusion of some of the large molecules which are present within the cell, and which may be part of the buffer system, it is occasionally necessary that there occur migrations of anions and cations between the intracellular and extracellular fluids coincident with buffering action. It is known, for example, that an increase of tension of carbon dioxide in the venous blood followed by the diffusion of carbon dioxide into the red blood cell would promote a diffusion of bicarbonate ions from the red cell to plasma water. In order to maintain electrical neutrality this movement is coupled with a shift of chloride from plasma to red cell water. It is also known that exchange of cations such as potassium, sodium and perhaps calcium for hydrogen ions across cell membranes, and between the extracellular fluid and bone, also plays a significant role among those factors which operate to maintain pH within narrow limits. The exact extent to which this occurs is not known but in certain experiments it has been estimated that about 50% of an acid load administered to dogs experimentally was neutralized by an exchange of hydrogen ions for sodium and potassium ions.

As was mentioned previously the bicarbonate CO_2 buffer system is highly efficient because of the capacity to remove CO_2 via the lungs. The mechanism which governs the amount of CO_2 lost by the lungs can also be construed as a feedback mechanism. The input to the system is the pCO_2 and pH of extracellular fluids, which are probably monitored by receptors located in the medullary respiratory centers of the central nervous system. These react in such a way that an increase in the retention of CO_2 or in the hydrogen ion concentration promotes an increase in pulmonary ventilation, and a decrease in these two variables inhibits ventilation. There are secondary receptors which monitor pH and pCO_2 as well. The respiratory activity is therefore modified in such a manner as to restore the initial deviation in pCO_2 or pH to normal, and thus complete the negative feedback loop. A period of respiration during which the pCO_2 is kept elevated appears to diminish the sensitivity of the respiratory centers to a specific level of pCO_2 , and, in contrast, a period of hypocapnea seems to enhance the sensitivity of the respiratory center to a given level of carbon dioxide tension thus changing open loop gain. It cannot be excluded of course, that prolonged elevations of CO_2 tension will be accompanied by an increase in the bicarbonate concentration in the plasma and thus change the pH of the body fluids, perhaps also effecting a change in the apparent sensitivity of the cells to CO_2 . As far as

acid base equilibrium is concerned, it should be mentioned that a number of nomograms which will correctly predict the ratio between the concentrations of various anions and cations in the plasma in relation to the pH and the carbon dioxide tension are available in the clinical literature.

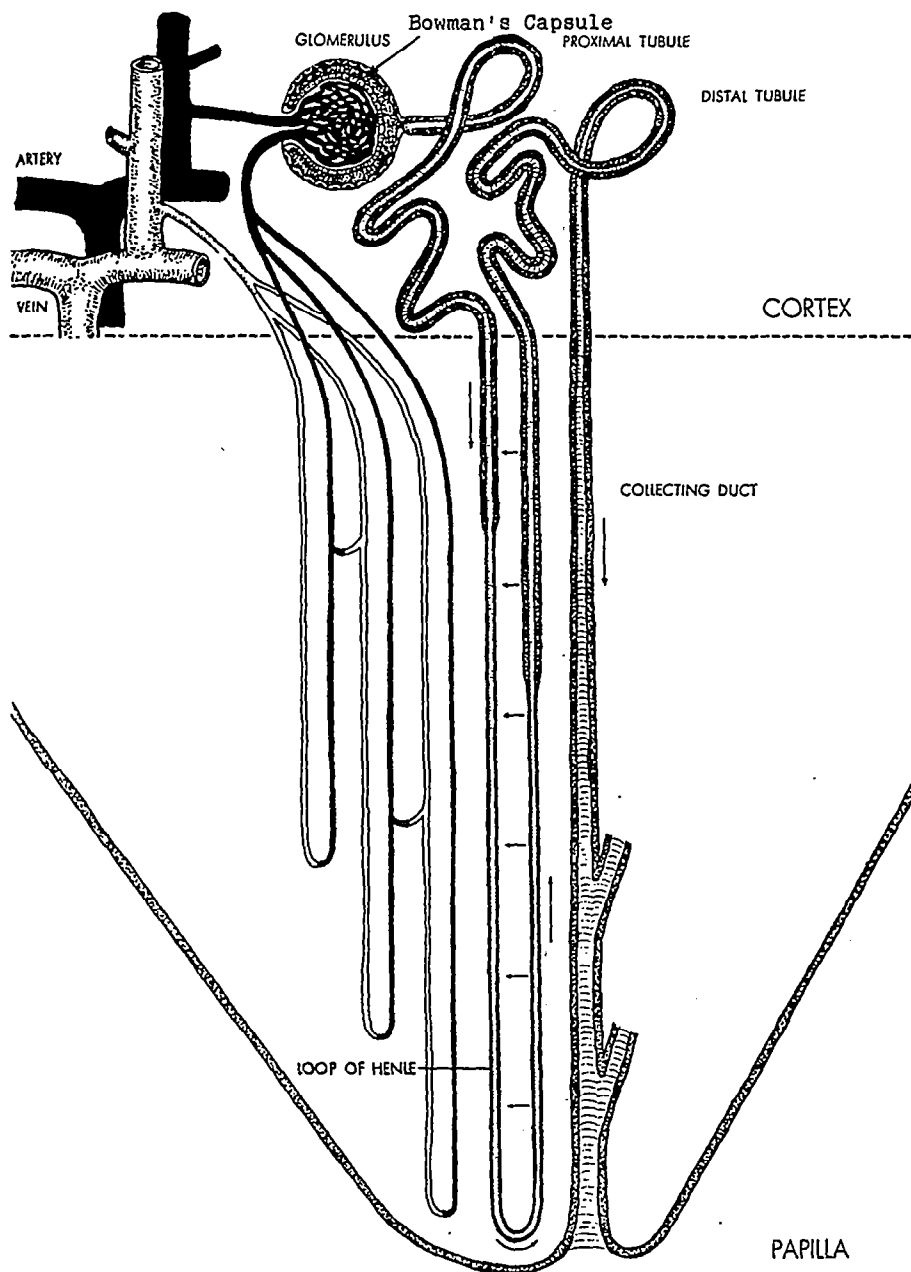
PHYSIOLOGY OF THE KIDNEY

Introduction

An understanding of the principles underlying renal physiology is crucial to a comprehension of the regulation of body fluids, since the kidney is the principal regulator of both volume and composition of the extracellular fluid, and, indirectly, the intracellular fluid. Within this very complex structure, a number of specific exchanges between intra- and extracellular fluids and between various components of the extracellular fluids and the urine are performed, many of which have no clear parallel to the regulation of the physiologic fluids in other tissues.

In this necessarily brief review, we will attempt to cover some of the basic areas of renal function, and some of the factors which are related to these functions.

The formation of urine. Each kidney is composed of a million or more similar units called nephrons, which function in parallel. Each nephron consists of a long tube with certain anatomical peculiarities at various points along its course. The afferent end of the tube is connected to a specialized structure called a glomerulus, the efferent end of which leads into collecting ducts which ultimately allow urine to course into the bladder. The nephron is diagramed below.



(P.F. Scholander. The wonderful net. Sc. Am. 196:105, 1957)

From the diagram one can visualize a sequence of events leading to the formation of urine somewhat as follows: The glomerulus consists of a number of capillaries carrying arterial blood, which are encased by the double walled structure of Bowman's capsule. Since the pressure in the glomerular capillaries is relatively high (75 mm of mercury) this hydrostatic pressure creates an outward gradient for all substances present in the plasma. Since this gradient is virtually unopposed in Bowman's capsule, transudation of all substances which are capable of passing through the double layer of semi-permeable membrane are filtered from the capillary into Bowman's capsule and thence into the tubule of the nephron. A detailed consideration of the nature of this membrane is beyond the scope of this discussion. It is only necessary to remember that this semi-permeable membrane does not allow the diffusion of larger molecules such as proteins, but almost all the other constituents of plasma, that is, most small non-ionized molecules such as glucose and urea, all anions and cations which are present in the plasma and a certain amount of plasma water all pass through the glomerular capillaries into Bowman's capsule. This process is termed glomerular filtration. The fluid which collects in Bowman's capsule, termed the glomerular filtrate, then begins its long journey through the proximal tubule, down into the medulla of the kidney, through the thin segment and through the thick segments of Henle's loop where in the medulla of the kidney it makes a sharp hairpin turn and returns to the vicinity of its own glomerulus in the cortex, passes through a

number of convolutions once more, (the "distal convoluted tubule"), and then passes into the collecting tubules which ultimately coalesce to form larger and larger ducts, finally ending in a structure called the renal papilla through which urine flows into the renal pelvis, thence into the ureter, thence into the bladder, and ultimately to the exterior. Since the glomerular filtrate is essentially an ultrafiltrate of plasma, concentrations of the various ionic species and other small non-ionized molecules are essentially identical to those concentrations in the plasma. In order for any significant alteration in the concentration of any ionic species or in the total osmolarity of the urine to occur, energy must be expended. In fact, all further manipulations which are performed upon the tubular fluid after filtration involve the expenditure of energy either at the tubular membrane, within the tubule cells themselves, or in the immediate vicinity of the tubule.

The blood supply to the renal tubules is principally derived for each nephron from the artery to its glomerulus, which after breaking up into a network of glomerular capillaries, coalesces again into an efferent arteriole, breaks up once more into capillaries which surround all of the tubular segments, subsequently combining again to form venules. So that in addition to operations performed upon the urine, the tubule cells probably perform additional functions upon the blood which circulates in their vicinity.

Methods of studying renal function. There are three principal ways in which renal function has been studied. The first is to simply regard the entire kidney as a "black box" which has as its input certain parameters relating to rates of flow and composition of plasma, and has as its output similar parameters relating to the volume and composition of urine. The second method involves the use of the "stop-flow" technique which by closing off one ureter allows urine to accumulate for long periods of time in all the glomeruli of that kidney. When the pressure in the tubules has built up to the point where glomerular filtration can no longer occur, the column remains static and whatever individual operations are performed by the various tubule cells are therefore intensified since they are continually performed upon the same given volume of fluid. These changes that have been produced by individual segments of the tubules can then be examined by releasing the column of fluid and examining different aliquots whose origin is presumably from different sites in the nephron. The third method involves the use of direct micro-puncture of various sites along the glomerulus followed by micro-chemical analysis of the fluid obtained.

These latter two methods have helped to clarify the nature of the processes which are sequentially performed upon tubular fluid in order to convert it to bladder urine.

If we consider the kidney as a black box which performs certain operations upon plasma which result in the formation of tubular urine, we may then form certain conceptions about how certain molecular species are handled by the kidney.

Let us assume that we have a substance which is easily filterable through the glomerulus and then passes without any alteration through the entire course of the tubule of the nephron and into the bladder urine. Also let us assume that during passage of this substance down the tubule none of it is removed by the tubule cells and no more of it is added by the tubule cells. (Such a substance is the hexose sugar inulin.) It would then follow, that if one were able to maintain a measurably constant plasma concentration of inulin by a continuous intravenous infusion of this substance, and then, if one were able to monitor the volume of the output of urine during a specified time interval during which plasma inulin is constant, and if one were able to determine the concentration of inulin in the excreted urine, it would then be feasible to determine the volume of plasma which had been filtered through the glomerulus during this specified time. By knowing the volume of urine excreted and the concentration of inulin in the urine, one then knows the total amount of inulin excreted. Since plasma inulin was held constant during this period it follows that

$$C = \frac{UV}{P}$$

where

C = number of millilitres of plasma "cleared" per unit time,

U = urine concentration of inulin,

V = volume of urine excreted during the unit time,

P = plasma concentration of inulin.

C is also referred to as the "clearance" of inulin, that is the number of millilitres of plasma from which inulin was removed or "cleared" in a given time. Because inulin is a substance which is only filtered but is not secreted or reabsorbed by the kidney tubules after filtration the clearance of inulin has been taken

to be equal to the glomerular filtration rate. This value, when determined in adults, is found to lie normally between 110 and 130 ml/min. This would imply that some 180 liters of plasma are filtered through the two million glomeruli of both kidneys per day.

From the above considerations, it would follow that if the clearance of a substance other than inulin would be determined to be greater than the glomeruli filtration rate, one must then conclude that additional amounts of this material were added from the plasma after filtration was accomplished, presumably by active transport from the peritubular capillaries into the lumen of the tubules. Conversely if the clearance of a substance were found to be less than the glomerular filtration rate, the conclusion must be that there was a net transfer of this substance from the tubular urine back into the plasma.

Some substances such as Diodrast and para-aminohippuric acid are extracted almost completely from the plasma both by glomerular filtration and by the tubule cells from the peri-tubular blood in one passage through the kidney. The clearance of either of these substances is used as an estimate of the total renal plasma flow.

The processes of removal or tubular reabsorption and tubular secretion (extraction of material from peri-tubular blood and transport into the tubular urine) are active rather than passive, that is they require expenditure of energy by the tubular

cells. Since most physiological energy transport systems have a limit to the rate at which they can function, it is to be expected that at certain very high concentrations of actively transported solutes (either in plasma or tubular fluid) the capacity of the tubular cell to transport these solutes will be exceeded. For example, under normal circumstances the concentration of plasma glucose is usually held constant between 100 and 120 milligrams per cent, and there is usually no detectable glucose in the urine since at this level it is usually completely removed by the tubular cells. However, if the concentration of glucose in the plasma is raised above 200 milligrams per cent, the amount of glucose which will be present in the glomerular filtrate will be greater and consequently the amount of glucose delivered to the renal tubules per unit of time will be greater than that which can be removed from the tubular urine by active transport, and thus some glucose will then pass into the bladder urine. The higher the plasma concentration of the glucose, the closer the clearance of glucose will approach the glomerular filtration rate since the fraction which is reabsorbed becomes progressively less in comparison with that which passes down the tubule.

Similarly, there is also a "tubular maximum" or T_m for substances which are secreted such as PAH, and when its concentration is raised sufficiently high a certain portion of the PAH will escape the process of tubular secretion. The clearance of PAH which initially was considerably higher than that of inulin

(or the GFR) will now approach this value asymptotically. With these foregoing considerations in mind, we can now turn to a consideration of the handling of some specific substances by the kidney.

Renal Excretion of Water

Water is generally considered to be freely permeable through all capillary and cell membranes. Along the renal tubules however, the situation is somewhat different; in order to regulate water excretion the permeability of the tubular epithelium to water is altered at several points. There is also evidence to suggest that permeability of the tubular epithelium to water is under the controlling influence of a hormone secreted by the posterior pituitary gland, antidiuretic hormone (ADH).

The capacity for water reabsorption from the renal tubules is enormous, since approximately 178 to 179 liters per day are usually reabsorbed, of the 180 liters that filter through the glomerulus. The entire sequence of mechanisms which operate on the glomerular filtrate to reduce its volume and alter its tonicity have not yet been completely elucidated, but enough data has been accumulated to conceive of the general sequence of operations as follows:

Water is generally thought to leave the tubular fluid by passive diffusion along concentration gradients which are established by the active transport of solutes. These concentration

gradients are maximized in the region of the loop of Henle because of the anatomic arrangement of the loop and the creation of a counter-current exchange process. (Vide infra).

The first phase of water reabsorption is responsible for about 85% of this reabsorption, and it transpires within the proximal convoluted tubule. Measurements of this region by micro-puncture studies have indicated that the proximal tubular fluid is isosmotic with plasma, suggesting that the passive diffusion of water along concentration gradients is so rapid that it can be considered to be directly absorbed along with solutes. The fluid, after passing out of the proximal convolutions then passes into the loop of Henle, down into the medulla and again into the cortex where it enters the distal tubule. The fluid in the proximal portion of the distal tubule is hypotonic with respect to plasma. It is thought that the loop of Henle is relatively impermeable to the passage of water (at least the ascending limb), and that active transport of solutes occurs in the loop of Henle into the interstitium surrounding this portion of the tubule.

The second phase in the reabsorption of water occurs in the distal convoluted tubule. Solutes are also reabsorbed to some extent in this portion of the tubule. The permeability of the distal tubule to water is thought to be regulated by the level of circulating antidiuretic hormone. In the absence of ADH, the fluid in this segment comes into osmotic equilibrium

with its environment and is isosmotic with plasma when it passes into the first portion of the collecting duct. With intermediate quantities of ADH, more and more water is reabsorbed up to the point at which the fluid becomes isosmotic. The epithelium of the collecting duct is also controlled with regard to its permeability to water by the presence of antidiuretic hormone. It should be recalled that the collecting duct travels through the medulla of the kidney in order to reach the renal papilla. It is known that as one approaches the renal papilla from the renal cortex, the interstitial fluid surrounding the collecting duct becomes progressively hyperosmotic with respect to plasma. Micropuncture studies of the fluid in the collecting duct when maximal activity of antidiuretic hormone is present have substantiated the fact that at all levels along the collecting ducts the fluid within it is isosmotic with respect to the surrounding interstitial fluid. Thus, as the urine approaches the renal papilla it is exposed to a hyperosmotic environment which allows further maximal water reabsorption.

The origins of this osmotic gradient which is present between the cortex and the renal papilla is thought to reside in the peculiar anatomical arrangement of the loop of Henle. It is thought that solutes are actively transported from the lumen of the ascending limb of the loop of Henle into the interstitial fluid; however, water does not diffuse freely across the ascending limb epithelium. The fluid in the upper portion of the ascending limb is thus made hypotonic with respect to plasma. On the other hand, the epithelium

of the descending limb of Henle's loop is permeable to solutes (and perhaps to water although this is not definitely known). Therefore, if isosmotic fluid is introduced into the descending loop of Henle's loop a counter-current exchange will be set up to operate so that as the fluid courses along the descending limb of Henle toward the medulla the fluid will become progressively more hyperosmotic with respect to plasma. After the lower portion of the loop is traversed and the fluid crosses along the ascending limb, it will become again progressively hypotonic, until the point where it enters the distal convoluted tubule, when it is hypotonic with respect to plasma. This sequence of events will cause the interstitium surrounding the loop of Henle to also assume an osmotic gradient which increases as one progresses from cortex to medulla. It is this environment which the collecting ducts must also traverse; and this environment provides the strong osmotic gradient necessary for maximal reabsorption of water.

It must be appreciated that the energy which is required to drive all phases of the process of water reabsorption is derived primarily from active transport of ionic solutes such as sodium ions. Antidiuretic hormone, although a major factor in the regulation of water excretion, acts merely as a valve to control the rate of passive diffusion of water along concentration gradients which had been previously established by the active transport processes.

The mechanism by which antidiuretic hormone alters the permeability of the membrane of the distal convoluted tubules and collecting tubule membranes is at present not clear.

Excretion of Sodium by the Kidney

The reabsorption of sodium by the kidney is a complex and poorly understood process. The kidney is extremely efficient in absorbing sodium since the equivalent of approximately 1.5 to 2 kilograms of sodium chloride are filtered through the glomerulus daily whereas under normal circumstances the amount of sodium in the urine is less than 1/1000 of this filtered load. Since the quantity excreted represents such a small fraction of this load, it is apparent that a small change in the rate of filtration of sodium (i.e. the glomerular filtration rate) could easily account for a two to threefold change in the daily excretion of sodium. Such changes in the glomerular filtration rate would be, in fact, too small to be significantly detectable by the usual means employed to measure the glomerular filtration rate. Thus, in studying the problem of renal disposition of sodium, one is faced with the difficulty of separating reabsorption of sodium from changes in filtration rate of sodium.

As far as can be determined, the tubular operations performed upon glomerular filtrate to reabsorb sodium are complex and are interrelated to some extent with the reabsorption and excretion of other ions, such as bicarbonate, chloride, hydrogen, and

potassium. The reabsorption of sodium is further conditioned by many other factors, including the volume and composition of body fluids, acid base equilibrium, the rate of excretion of potassium, the rate of excretion of osmotically active solutes and activity of certain hormones such as aldosterone and, perhaps, antidiuretic hormone.

The specific mechanisms involved in the reabsorption of sodium are unclear but a general scheme can be outline as follows:

The largest fraction of sodium reabsorption occurs in the proximal convoluted tubule and almost certainly involves active transport. It is known that sodium can be transported across the lumen of the proximal tubule both against electrical gradients and also against concentration gradients. It is possible that the transport may be partly passive and facilitated by the colloid osmotic pressure of the peri-tubular capillaries, since there is no significant hydrostatic pressure gradient across the tubule under normal conditions. The reabsorption of sodium in the proximal convoluted tubule is limited by a concentration gradient such that the luminal concentration usually does not fall below about 60% of that of the plasma. In order to maintain electrical neutrality across this portion of the tubule, sodium must either be exchanged with another cation which flows into the tubular urine or else reabsorbed together with the chloride ion. The latter process is thought to be predominant in the proximal convoluted tubule. The reabsorption of chloride at this site is very likely

passive and probably due to the electrical gradient established by the active transport of sodium. Active transport of chloride, however, cannot be excluded as occurring in this segment of the nephron

The second and more important phase of sodium transport occurs in the distal convoluted tubule. There are several differences between this reabsorption and that which occurs within the proximal tubule. First, a greater fraction of reabsorbed sodium is transported in exchange with hydrogen and potassium ions rather than with an accompanying reabsorption of chloride. Sodium, under certain circumstances, is also exchanged with ammonium ion. Second, there is no limiting concentration gradient in this portion of the nephron and the concentration of sodium in urine in the distal convoluted tubule may fall very near zero.

Under certain circumstances where a large amount of water is not allowed to be reabsorbed in the proximal tubule, a considerable amount of sodium may escape reabsorption at this site. (It is possible to obtain such a situation by infusing large amounts of inulin intravenously. Since this substance cannot diffuse freely across the tubular membranes much of the water which carries this solute is not reabsorbed since proximal tubular fluid cannot be hyperosmotic with respect to plasma. In such a case the flow rate of tubular urine remains high and other solutes such as sodium then escape absorption at this proximal site.) As a consequence a larger volume of fluid courses through the remaining segments of the nephron. Although the distal tubular mechanisms

for the reabsorption of sodium may accelerate, they are still usually inadequate to cope with the large quantities of this cation that gain access to this portion of the nephron, and hence a variable amount escapes reabsorption even at the distal site and is allowed to enter the bladder urine.

The volume receptor and sodium excretion. The ingestion and excretion of water under normal circumstances are the final regulators of the volume of all body fluids. It is important to recognize, however, that the principal stimulus to water excretion and intake is the concentration of osmotically active solutes, and that a relative deficit of water (such as provoked by an increase in solute concentration) will provoke the same mechanisms for water regulation as will an absolute deficit of water (in the presence of the same amount of solute). It is then possible, for example, to ingest a very large volume of isotonic solution and if no elements were available for absolute volume regulation, this would not be accompanied by any stimulation to excretion of water. One mechanism for the regulation of the volume of circulating plasma has been previously alluded to in connection with antidiuretic hormone. There is, however, one other feedback mechanism which is concerned with regulation of the volume of extracellular and intracellular fluid by regulation of the excretion rate of sodium. Since sodium is the principal cation of the extracellular fluid and is the most osmotically active molecule, it follows that regulation of the concentration of this ion in extracellular fluid will be followed by appropriate

changes in total volume of the extracellular fluids, if tonicity is to be held constant.

The manner in which changes in total volume of the extracellular fluids can be detected is not known at the present time. Among other speculations, it is felt that an increase in extracellular fluid volume somehow influences renal hemodynamics. It is possible that the transducer which is sensitive to changes in blood volume is located in a small collection of specialized cells located along the afferent arteriole to the glomerulus, (the "juxtaglomerular apparatus"). A number of lines of experimental evidence have emerged to support the following scheme:

An increase in total fluid volume of the body is detected in some manner by the juxtaglomerular cell. These cells then elaborate an enzyme ("renin") which acts to catalyze the formation of a polypeptide from constituents present in the plasma. This polypeptide or group of polypeptides known as angiotensin then acts upon the outermost zone of the adrenal cortex (the "zona glomerulosa") to stimulate the elaboration of a substance called aldosterone. Aldosterone then acts upon the distal tubule of the nephron to enhance the reabsorption of sodium at the same time increasing the excretion of potassium. These effects on sodium reabsorption and potassium excretion are closely related, and if, for example, the distal tubular mechanism is already acting to absorb all of the sodium in the tubular fluid, administration of aldosterone will not enhance potassium excretion.

We thus have a feedback mechanism in which contraction of the total fluid volume is sensed in some manner by transducers in the kidney, so that the serial formation of a number of substances is then initiated, culminating in the increased secretion of circulating aldosterone, which then exerts its effects upon the distal tubular mechanism to enhance sodium reabsorption, thus ultimately increasing the total volume of body fluids and completing the feedback loop.

The Renal Excretion of Potassium

On the basis of a number of clearance studies which have been performed using potassium, it is now known that the renal handling of potassium involves glomerular filtration, almost complete reabsorption, and finally, secretion of this ion into the distal tubular fluid. The secretion of potassium is now thought to involve an ion exchange mechanism whereby one sodium ion is reabsorbed for every potassium ion that is excreted. Furthermore, potassium is not the only ion, which is thought to exchange with sodium, that is reabsorbed at the site of the distal tubule; hydrogen ions are also secreted into the tubular lumen in exchange for sodium. Whether potassium or hydrogen ion is predominantly excreted in exchange for reabsorbed sodium depends to some extent upon the relative availabilities of each of these ions. Thus, for example, if for some reason the delivery of sodium ions from the proximal tube is diminished, or if distal reabsorption of sodium is reduced because of a low level of circulating

aldosterone or other adrenal cortical hormones, the secretion of potassium will be compromised. Furthermore, if the concentration of hydrogen ions within the renal tubular cells is significantly increased, these ions will then be in a more favorable position to compete with potassium for exchange with sodium ions that are being reabsorbed and thus again the tubular secretion of potassium will be somewhat inhibited. This renal mechanism for excretion of potassium would imply that under normal circumstances, the urine is not free of potassium; the concentration of potassium in the urine does not usually fall below that of the plasma.

Excretion of Hydrogen Ions by the Kidney

As has been alluded to earlier, the bicarbonate buffer system, operating together with carbon dioxide elimination by the lungs acts very efficiently to maintain the pH of the plasma within certain specified limits. However, efficient operation of this buffer depends to a large extent upon stable concentrations of other anion and cations within the plasma, whose rates of excretion are regulated in the kidney. The kidney also significantly affects the pH of the body fluids by altering the net rate of excretion of hydrogen ion. Some of the poorly dissociated acids may be excreted intact, but their excretion is limited since the lowest pH of the urine that has been observed is about 4.5, and at that level only minor concentrations of even weak acids can exist as such. In order then for large amounts of hydrogen ion to be excreted it is necessary that the urine be adequately buffered or that the

hydrogen ion be coupled in a poorly dissociated form to a base.

The present concept regarding handling of hydrogen ions by the kidney is that hydrogen ions are actively secreted into the tubular fluid in exchange for sodium ion reabsorption. The source of hydrogen ion is thought to be carbonic acid which is present within the tubular cell, which in turn is derived from the prior combination of carbon dioxide and water. The carbon dioxide may have been present due to diffusion from the plasma, or it may be the end result of the metabolic activity of the tubular cell. The rate of production of hydrogen ions would be slow if it depended upon the uncatalyzed hydration of carbon dioxide; but the cortex of the kidney is rich in an enzyme, carbonic anhydrase, which permits rapid hydration of carbon dioxide to carbonic acid, which may then dissociate into hydrogen and bicarbonate ions. The sequence of events then for the secretion of hydrogen into the tubular fluid would be as follows:

Carbon dioxide is present in the cell. It combines with water and is catalyzed by the enzyme carbonic anhydrase to form carbonic acid. This in turn dissociates into hydrogen ion and bicarbonate ion. The hydrogen ion is actively transported into the tubular fluid and is exchanged for a sodium ion. The reabsorbed sodium ion together with the bicarbonate ion then diffuses into the interstitial fluid and ultimately to plasma. The hydrogen ion which is secreted into the tubular fluid may be handled in one of several ways. First, in addition to the

bicarbonate ion which is formed in the tubular cell by the mechanism outlined above, there is also bicarbonate ion present in the plasma which is filtered through the glomerulus and is present in the tubular fluid. The secreted hydrogen ion will then combine with this bicarbonate ion to again form carbonic acid which will ultimately dissociate to form carbon dioxide (which is reabsorbed) and water (which continues down the tubule until it reaches the bladder). Alternatively, the secreted hydrogen ion may combine with other buffers such as sodium diphosphate to form the monophosphate ion which is excreted as such. These two substances act as buffers to keep the pH of the urine above the level of 4.5.

Another mechanism for the protection of the pH of the lumenal fluid is the secretion of ammonia. Ammonia is formed in the tubular cell by deamination of certain amino acids, among them glutamine. The rate of production of ammonia is accelerated by a decreased availability of buffers in the tubular fluid and a lowering of the pH of this tubular fluid. The mechanisms responsible for initiating the increase in the production of ammonia are not known; however it is thought that free ammonia is formed in these cells and that the ammonia then diffuses across the tubular epithelium into the lumen where it combines with hydrogen ion to form ammonium ion and thus raises the pH of the tubular fluid.

It will be noted that the net effect of the reactions involving the excretion of hydrogen ion include the reabsorption of sodium, and the reabsorption of bicarbonate; thus, for each hydrogen ion which is exchanged for sodium, one molecule of sodium bicarbonate is returned to the interstitial fluid.

As was mentioned earlier the reabsorption of sodium can be accompanied by either an ion exchange with hydrogen or potassium ions, or by an equivalent reabsorption of a negative ion, the chloride ion, to preserve electrical neutrality. If an ion such as sulfate, which is poorly permeable across the membrane, is infused with sodium, more hydrogen ion will be exchanged for sodium because of the diminished availability of the anion for transport along with the sodium.

Certain other factors also operate to control the rate of excretion of hydrogen ions. Decreased pH within the tubular cell or increased CO_2 tension in the plasma, both of which could readily make more hydrogen ions available for secretion, in fact accelerate the rate of secretion of hydrogen ions. A diminished concentration of potassium in renal tubular cells by favoring exchange of sodium with hydrogen ions also promotes acidification of the urine. Finally, there appears to be a limiting concentration gradient against which hydrogen ion cannot be

secreted into the tubular fluid, and this gradient is approximately 800:1, which would explain why the pH of the urine usually does not fall below 4.5. If the concentration of buffer substances in the tubular urine is high, again the rate of hydrogen ion excretion will be favored.

2. CONSIDERATIONS OF SOME ASPECTS OF CONTROL SYSTEMS OPERATING TO REGULATE THE VOLUME AND COMPOSITION OF BODY FLUIDS

Introduction

At the present time there is substantially no science of control systems existant in the field of renal physiology. In recent years, physiologists have come to appreciate that in order for homeostasis with regard to body fluid volume and composition to exist, a number of feedback and self-regulatory mechanisms must be in continuous operation. Beyond this glimmering of light, however, there exists no theoretical or experimental body of data whose purpose is to better define the control systems which educated intuition says must exist.

This is, in essence, what we must approach. Without a consideration of the control systems involved in fluid balance it is impossible to arrive at even an approximate philosophy of design, such that one would be able to predict the behavior of the system. Nor can one, from first principles, derive a design philosophy for the kidney or other body fluid regulators. It is necessary not only to be thoroughly familiar with the large mass of data which has been accumulated in the field of regulation of body fluids, but also to be thoroughly cognizant of the many difficulties involved in applying scientific techniques to the consideration

and experimental investigation of any complex biological system.

Biological Control Systems: Definitions and Restrictions

A straightforward application of servomechanism theory to body fluid regulation encounters formidable problems. A biological control system (let us say for example, the mechanism of thirst) consists generally of an input to a sensing device, whose information is then processed by an amplifier which produces a function of the input signal which in some fashion acts to alter the input. The sensing device or sensitive cells may be located adjacent to the amplifier or may be located at a distance from it. The input to the sensor is generally contaminated by noise. The output of the amplifier is usually a non-linear function of the input signal. Furthermore, the output signal of the amplifier may be acted upon by another control system before it actually exerts its effect upon the original input. In the control system which regulates thirst, the input signal is a contraction of the extracellular fluid volume or a change in osmolarity of the extracellular fluid. The sensing devices are sensitive to changes in osmolarity and an increase in osmolarity is processed by the amplifier (this is a hypothetical structure probably located in the diencephalon) which yields the sensation of thirst as an output function of the change in

osmolarity. This sensation of thirst ultimately causes the ingestion of large amounts of water which tend to expand the extracellular volume and return the osmolarity to its previous level, so that the sensor and amplifier system can again stimulate the thirst sensation. A similar analogy can be made for other systems which act to regulate volume and composition of the body fluids; these will be considered below in some detail. Several comments should be made about each of the components in a biological amplifier, in order to understand how they may differ from conventional control systems.

The amplifier. This is defined as that portion of the system which lies between the sensor and the output of the system. Amplifiers which are concerned with the regulation of body fluids have one peculiar and significant characteristic: the proper functioning of the amplifier is itself dependent on maintenance of its own homeostatic environment, since the amplifier is contained within the environment which it regulates; it thus receives a signal from this environment and produces an output which acts in some fashion upon the environment in general, but also on its own environment in particular. It may do this by generating a function which will act upon some component outside its own immediate environment which will then act upon its environment, or it may regulate the environment

without resorting to outside systems (see Fig. 1). Such a system has a tendency towards instability in that certain inputs may change the environment sufficiently so that the mode of operation of the amplifier is changed. Although within normal or tolerable ranges of values for the environment, an amplifier usually functions in a time invariant mode, this is not always the case, as we shall see.

Linearity of biological amplifiers. It is necessary to consider linearity of biological amplifiers as compared to conventional amplifiers, with certain reservations. The first reservation is the one mentioned above, that most biological amplifiers, being tied into their environments can become markedly non-linear when this environment is changed sufficiently. However, even within the normal ranges all of the amplifiers that we are going to discuss are non-linear for two reasons: First, whereas the input signal to these amplifiers is usually an absolute value (such as sodium or potassium concentrations, osmolarity, pH, CO₂ tension) the output of the amplifier usually acts to change rates of flow or transfer of various materials from one compartment to another. The rate and concentrations are non-linearly related. In first order chemical processes the rate is related to the logarithm of the concentration; a linear dynamic characteristic is present only for small signals.

A much more difficult problem revolves about the fact that most biological amplifiers operate in close relationship

to another control system, which monitors either the same or another variable; thus the closed loop characteristics of the system as a whole become exceedingly complex and the dynamic characteristics of each of the individual amplifiers may have little if any physiological significance.

It must also be mentioned that the actual process of accumulating data regarding these systems is not a simple matter. Since the mode of operation of a number of systems may change with time (as we shall see) it becomes very difficult to introduce to any one system a reproducible signal that has physiological significance. The actual measurements may disturb the system to such a degree that one is not observing the physiological operation of the system. For example, if one were to directly operate upon the animal in order to introduce catheters for measuring arterial and venous sodium concentration, one would have to exclude the possibility that surgical intervention, by changing the secretion of adrenal cortical hormone has changed the pattern of sodium excretion. Similar problems exist for almost every biological measurement.

Signal: Although what constitutes a biological signal, that is the input to the biological amplifier, is usually a matter of definition, certain systems have somewhat complex signals in that the input presented to the sensor may be a complex function of the actual variable which is to be monitored by the system. For example, in the thirst control

system referred to above, the output of the amplifier (thirst) regulates the amount of water ingestion and thus, influences the total volume of body fluids. The variable monitored is extracellular fluid osmolarity which is closely related to the concentration of sodium in the extracellular fluid. Although in most situations a contraction of extracellular volume usually occurs as a loss of water without a loss of salt, this is not invariably the case; and it is quite possible for a large (isomotic) decrease in volume to occur with no stimulus presented to the osmo-receptor. In addition, the osmo-receptor is also to a degree dependent upon the constant functioning of the various mechanisms which are in operation to control the concentration of sodium in the extracellular fluid; these too are largely dependent upon the total volume of extracellular fluids. Thus, although the osmo-receptor acts to control extracellular volume, it does this only in an extremely indirect fashion, and depends for its stability upon an intact mechanism for controlling the rate of excretion and ingestion of sodium. How one defines a biological signal in this system is not of much importance, but it should be recognized that any definition thus applied is valid only within the context of a properly functioning multiplicity of regulatory systems.

Noise: The definition of noise again depends largely upon the definition of what constitutes the signal. If, in the example above, one would consider that the osmo-receptor

is actually monitoring total fluid volume, then one would have to say that a large component of noise is introduced by various other systems which regulate the concentrations of sodium and of water. Such a definition, however, contributes little insight into the actual mechanism of the thirst control system.

If one rather considers that osmolarity is, in fact, the variable being monitored and that this variable reflects only partially the total fluid volume, then noise can be defined as those components of the signal which would tend to generate an error in the accurate determination of the osmolarity of the extracellular fluid.

In general, there is both an extrinsic and intrinsic noise component to biological systems. Intrinsic noise can be generated within the sensor itself as a result of the alteration of its own immediate environment. It may arise as the result of external stresses placed upon the sensor and amplifier such as hypothyroidism or fever where the sensitivity of the sensor may change, or the mode of the system may change. Noise may also be generated within the system itself; cascading of one system upon another will multiply the noise. Within the system, noise may be generated by the central nervous system and may be manifested as a time lag (e.g. delay from the time of experiencing the thirst sensation to the time of drinking water), a non-recognition of the sensation of thirst

(due to other pre-occupations) and other influences.

Extrinsic noise may arise from several sources. For example, a high concentration of urea in the blood stream and thus an increase in total blood osmolarity will not alter the osmolarity sensor since urea is generally freely permeable across the cell membranes, thus osmolar control may be deranged in pathologic conditions such as uremia.

In biological systems in general, a rough estimate of the magnitude of the noise level may be made by holding a given signal constant for a long period of time, assigning a measure to the "random" fluctuations in output occurring over a period of time, and then comparing these fluctuations with those that would be produced by appropriate variations of the signal. In practice this is an extremely tedious procedure except for the simplest of control systems, and has almost never been done, studies having been confined to "spot" measurements alone.

Sensors: Biological sensors involved in the control and regulation of body fluids monitor concentration levels rather than rates of flow or other variables. By what means the sensor is able to maintain a standard or reference value for the input from which deviations may be measured, is not clear. The assumption has almost always been that the sensor output signal is related to the magnitude of the error, but it has not been excluded that the output may be some function

of the derivative of the error. In most systems, integration of the error at the level of the sensor usually does not occur. Smoothing of an osmolarity signal is produced not in the sensor, but as a result of the fact that changes in the osmolarity of the extracellular fluid are transferred to the intracellular compartment, a large reserve of body fluids in osmotic equilibrium with the extracellular fluid. Thus, an abrupt change in osmolarity of the extracellular fluid would tend to be smoothed because of the large volume of fluid with which it equilibrates.

Lag: In the various control systems which regulate body fluids, although the individual lag of each sensor amplifier system may be brief, the total lag of any one control system is usually great enough so that an unstable, oscillatory mode may conceivably occur.

In addition, phase lags may often be created in a system because of the presence of two parallel biological control systems with different time constants. In such a case it can be seen that if a positive error input will pass through both systems, the output from the system with the longer time constant may arrive at a time when the system with the shorter constant has already returned the positive error back to zero, and thus an overshoot will be caused. Although a system such as this might be expected to produce oscillatory instability, no biological data exists for the systems which

regulate body fluids to show that such situations do, in fact, occur. Usually the system will go to its extreme value.

Sampling: In general, the data that has already been gathered concerning the regulation of body fluids would suggest that the sensors of the various systems involved are operating continuously. The frequency response of the biological sensors involved in this system cannot be easily evaluated because of the heavy damping effect of the various compartments which are closely linked to the extracellular fluid compartment. It is unlikely, however, even in pathologic situations that the sensor would be altered to such a degree that a change in its frequency response would cause instability in the system. Under such circumstances the sensor will actually cease to function (perhaps this can be construed as a reduction in frequency response).

For these reasons then, determining the dynamic characteristics of the sensor becomes essentially an experimental problem where one must be able to drastically and rapidly alter the vicinity of the environment surrounding the sensor. The physiologic reactance that is involved in this procedure is enormous and with our present techniques meaningful data on this point cannot be evaluated. For this reason also, although one would assume intuitively that the sensor does operate over a definite range of error this point is also difficult to evaluate.

Stability: At present no useful experimental techniques exist for evaluating the actual stability of any body fluid regulating mechanism. Empirically we know that: (1) despite the high sensitivity of all chemo-receptors and volume-receptors involved in regulation of body fluids and despite the long lags which are inherent in the multiplicity of systems necessary to regulate a single variable, there are only a few pathologic conditions in which the control system exhibits instability. No instance of oscillatory instability is known as yet. However, there is no reason to think that an oscillatory instability cannot be made to arise experimentally. For example, if one would infuse a large volume of hypertonic salt solution intravenously, the osmo-receptor would be triggered and anti-diuretic hormone would cause more water to be retained. Sodium secretion, however, would not be increased because of the lack of expansion of the plasma volume. If there were pathological hyperreactivity both of the sensitivity of the center for thirst to the osmo-receptor and to the kidney's mechanism for sodium excretion to aldosterone, then the increase of osmolarity set up by the infusion of sodium would provoke a hyperreaction in the form of exaggerated water intake which would thus dilute the extracellular volume and also increase its total volume, thus causing renal loss of sodium in large amounts which would in turn provoke a powerful water-retaining response and cessation of thirst for a long period of time.

This example should indicate several reasons why the fluid regulatory mechanisms are relatively resistant to oscillatory instability. These would include, first, the tendency of receptors to become less sensitive rather than more sensitive when they are pathologically altered, so that the gain of the system is rarely increased; second, that all of the systems that have been experimentally investigated operate continuously and proportionately, rather than on a "bang-bang" basis. Of course, these considerations do not preclude the possibility of non-oscillatory instability and, in fact, this is encountered frequently in the systems we are considering.

It is clear from the foregoing discussion that many of the concepts which have been developed in control system engineering are applicable to biological systems, if one recognizes the limitations imposed upon these concepts by the system. Furthermore, one of the main objectives of applying control system theory to biological systems is the visualization of the design philosophy of the system itself, which is a problem different from predicting the behavior of a system which conforms to an a priori design philosophy.

With these considerations in mind, we can briefly examine some of the simpler control systems operating to regulate the volume and composition of body fluids and perhaps gain some insight into some of their features and for their characterization and simulation.

Volume Regulation

In Fig. 2 we have a diagram which indicates some of the control system pathways operating in the regulation of the total volume of body fluids. From the biological standpoint, the concentration of water alone in body fluids, or the absolute body fluid volume of water, is neither monitored nor measured. Rather the body fluids are considered from two standpoints by the various sensors involved (1) the total volume of extracellular fluid without regard to solute concentration (a reflection of the intracellular volume of water) and (2) solute concentration (osmolarity of the body fluids).

There are at least four separate systems which regulate the total volume of body fluids, which we will refer to as loops I, II, III and IV. Loop I is a system which has as its input the osmolarity of body fluids, sensed by a group of cells in the hypothalamus, the osmo-receptor cells. This input is then operated upon by other cells in the diencephalon which elicit an output which consists of the sensation of thirst which ultimately drives drinking behavior. The time constant of this system is relatively long, usually lagging the change in osmolarity by several hours or days. The system can be, at least at the present time, considered to be approximately linear, since an increase in osmolarity of extracellular fluid will promote an approximately proportional increase in the sensation of thirst. In as much as the

intensity both of the sensation of thirst and the osmotic stimulus are both subject to biological limitations, there are limits beyond which the system is no longer linear. Death of the organism usually ensues before these limits are reached. Although the time constant of the system is long, the possibility exists that the thirst sensation is immediately shut off after the osmolarity of the body fluids has been returned to normal. The system may thus possess two modes of operation, a rapid mode for turning off the thirst stimulus and a slow mode for generating the thirst stimulus. Such a combination of systems would make for increased stability and tend to decrease the amount of overshoot. It is known for example, in diabetes insipidus where the sensation of thirst is so powerful that it may cause the drinking of 50 to 60 litres of water a day, that the administration of antidiuretic hormone, thus stopping the excretion of water, is followed by a cessation of the sensation of thirst as soon as the tonicity of body fluids approaches normal.

The mechanism by which the osmo-receptor is able to detect changes in osmolarity of body fluids is not known. It is possible that the actual volume of the cells may change as a result of water entering or leaving the cell to maintain osmotic equilibrium. The fact that the system operates on absolute values of osmolarity rather than the derivative of absolute values would suggest that these cells

have some reference value with which input osmolarity is being continuously compared; this reference value may be in the form of the initial volume of the cell, or may be concentration of a particular solute which cannot leave the cell. It is probably not related to membrane potentials since membrane potentials would not be expected to be linear with changes in osmolarity of fluid surrounding the osmo-receptor.

Feedback loop II has never been characterized as to its receptor or its amplifier. The existence of this loop is known only from experiments in which salt-depleted animals will tend to select food which has the greater concentration of salt. As the receptor and amplifier of this system are not known no statement can be made regarding linearity or stability of this system. The system probably does have a rather long time constant, probably even longer than that existing for thirst, but whether it operates in one or two modes is certainly not clear.

Loop III, which also operates through the osmo-receptor, has a rather rapid time constant. The same osmotic stimulus which will drive the output for thirst also drives cells in the pituitary gland to secrete the antidiuretic hormone, which in turn acts upon the collecting ducts of the kidney to reduce water excretion by affecting permeability of the collecting tubular cells to water. The reduction of excretion of water can begin within thirty minutes after

the osmotic stimulus is received by the osmo-receptor; in this sense then the system has a rather rapid time constant and acts to prevent loss of water when the organism is already volume-depleted. Although this system is intrinsically stable, under certain pathologic conditions, particularly in certain brain tumors and perhaps in carcinoma of the lung, a large component of noise may be inserted into the system either before or after the osmo-receptor, so that ADH may be either over- or undersecreted. If ADH were oversecreted in the presence of normal osmolarity, water would not be excreted and the total fluid volume of the body would tend to expand, accompanied by an increased total body sodium. Under these circumstances then, the secretion of antidiuretic hormone is a significant link in the maintenance of fluid volume, although in the final analysis, the osmolarity of the fluid would not change. A pathologic sensation of thirst or pathologic water drinking behavior usually will not disturb body fluid equilibrium greatly, since this will be compensated for by the remaining three control systems. But an error in the regulation of antidiuretic hormone can cause malfunctioning in systems II and IV.

Loop IV is the one which most closely monitors the total volume of body fluids, although in an indirect manner. The total volume of extracellular fluid is monitored by a poorly defined "volume receptor" which is known to be sensitive to changes in isotonic volume of the extracellular

fluid. The exact site of this receptor is not known. It may be located in the kidney or perhaps in the venous circulation of the bowel and other viscera. In the kidney it may monitor the total volume of extracellular fluid by changes in arteriolar or capillary pressures due to extracellular volume expansion. The effect of an expansion of extracellular volume is the eventual elaboration of aldosterone by the zona glomerulosa of the adrenal gland. The connection between the sensor and the amplifier which produces the aldosterone is not yet defined, but the possibility is that this is mediated by a hormone **which is extractable from the kidney, (renin)**. Aldosterone acts most probably on the distal convoluted tubule to enhance the distal reabsorption of sodium although it may have an effect on the proximal tubule as well. The net effect is a marked reduction of the amount of sodium which passes into the urine. The resultant increase in osmolarity triggers the osmo-receptor resulting in increased intake of water. The time constant of this system is probably longer than that of antidiuretic hormone (Loop III) but it is a relatively rapid system. This system is also approximately linear, within the normal range, but under certain pathologic conditions may become nonlinear. For example, in congestive heart failure, decreased blood flow to the kidney may trigger the secretion of aldosterone in large amounts causing a further over expansion in extracellular volume. Under these

conditions the gain of the system has been so reduced that the system can no longer operate to augment sodium excretion, and therefore tends toward an "infinite" increase in extracellular volume. A similar condition may occur in cirrhosis of the liver where the level of circulating aldosterone may be in part regulated by its rapid metabolism by liver cells. If the liver cells are injured so that they cannot metabolize and thus remove circulating aldosterone from the circulation, the level of aldosterone may rise independently of changes in extracellular volume. In this situation the positive gain of the system becomes so large that the system will again tend toward an infinite expansion of body fluids.

Briefly summarizing some aspects of volume regulation, it can be said that four control systems are in operation here: two systems which regulate the intake of water and salt, both with long time constants but with rapid shut off systems; and two systems with shorter time constants, both operating to inhibit the excretion of water and sodium. It is probable that each of these systems is continuous and since there is no contradictory data extant at the present time, it is approximately linear. Interruption of either of the longer-acting control mechanisms probably will not result in instability of the system. Interruption of either of the shorter acting feedback loops will usually result in a limiting or extreme output.

Handling of Electrolytes

In the present state of our knowledge concerning the regulation of the concentration of electrolytes in body fluid, it is impossible to draw a diagram of mechanisms operating to control serum sodium concentration, total body sodium, or the concentrations of other electrolytes. We can, however, offer a few conjectures about the nature of this regulation.

Sodium regulation. One of the factors which is operating to control the level of total body sodium is feedback loop IV (Fig. 2) involving volume regulation as was discussed above. It is obvious that if the serum sodium drops below normal the osmo-receptor system would tend to correct this deficit by restricting water intake and increasing renal excretion of water. The subsequent drop in extracellular fluid volume due to the excretion of water would tend to cause a rise in aldosterone level and renal reabsorption of sodium. Thus in one sense although the aldosterone pathway is used for the regulation of volume, indirectly it also serves as a regulator of the serum sodium concentration. Whether or not the concentration of the serum sodium is regulated in a more direct manner by sensors in extracellular fluid has never been demonstrated; particularly changes in serum concentration without changes in volume have not been shown to be productive of changes in sodium reabsorption or excretion. It is possible that prolonged deprivation

of sodium may cause an increase in the desire to ingest sodium, (pathway II, Fig. 2), as speculated in the earlier paragraphs. One very likely possibility is that the concentration of sodium in the extracellular fluid is not monitored at all; rather the concentration ratio of sodium across the cell membrane of the individual cells may be in fact the variable which is monitored and controlled. Slight changes in sodium concentration will effect membrane potential and perhaps as a partial consequence of this change or coincident with it, the rate of transport of sodium from the cell (by active transport processes) may change to keep the ratio of extracellular to intracellular sodium constant. Thus, the reference value of sodium concentration may actually be set by the cells themselves; variations in serum sodium may be compensated for by the change in renal excretion which is purely dependent upon the serum sodium concentration, and by the actions of the individual tubule cells which are interspersed between the urinary and extracellular environment. The existence of an integrated mechanism, which is not simply the sum of the actions of many individual, or specialized cells has never been shown. Furthermore, the interdependence of sodium reabsorption and potassium and hydrogen ion excretion may simply reflect the activities of many individual cells, each of which is specialized to perform the entire task (within the kidney) of electrolyte regulation without any coordinating influence upon them. As was mentioned in the introductory section, the operation of a counter-current system within

the kidney allows significant magnification of the effects of sodium transport by an individual group of cells, by repetitive performance of the same transport operation on one volume of tubular fluid.

Hydrogen ion excretion: If such is the case than the design philosophy of the renal mechanisms which operate to maintain stable electrolyte concentrations may in fact be the same as that of single cell which is attempting to maintain its own environment. As an example of this let us consider the excretion of hydrogen ions. It is known that increase in circulating hydrogen ions will result in an increase in the excretion of titratable acid by the kidney. However, it is probable that the kidney as such possesses no specific receptor for hydrogen ions, but that the excretion of hydrogen ions by the renal tubule cells is dependent upon the maintenance of electrical neutrality across the cell membrane of the tubule cells. There is no specific chemical mechanism in the kidney for the transport of plasma hydrogen ions; rather this is accomplished by the breakdown of circulating CO_2 into hydrogen ions and bicarbonate ions after the addition of water through the action of the catalyst carbonic anhydrase. Although hydronium ions are thus allowed to diffuse into the tubular fluid, these are not ions which have been removed from the extracellular fluid. Rather, the neutralization of excess acid in the extracellular fluid is accomplished by the resorption of a bicarbonate ion which is

generated concomitantly with the hydrogen ion. Since carbon dioxide appears to be freely diffusible into the tubule cells as well as bicarbonate and hydrogen ions, it would seem then that this mechanism appears to be passive and not under any particular regulation. As far as we know, the level of carbonic anhydrase is not a variable that is being monitored. The actual regulation of hydrogen ion excretion may come about as the result of the necessity for maintaining electrical neutrality across the tubular cell membrane. Thus for every hydrogen ion which diffuses out of the tubular cell, it is necessary that either an anion pass out of the cell along with it, or that one sodium ion be reabsorbed. It appears then, that most of the other ionic movements across the cell membrane may be considered at least at the present time to be moving along passive electrical and chemical gradients which are partially under the control of the active transport of sodium, and partially under the control of other influences which change their concentrations in the extracellular fluid.

Ammonia: There is not yet enough known about the renal excretion of ammonia to clarify its action in acid base equilibrium of body fluids. Since this is probably an active mechanism, it must be determined whether or not there is a specific pH receptor which accelerates the production of glutamine. If, for example, glutamine production is increased in nitrate induced alkyllosis, then one would suspect that the elaboration of glutamine in response to changes in pH is only

an apparent but not a real mechanism. Furthermore, the time constant of glutamine deamination and ammonia excretion is rather long (48 hours) for a system which is not mediated through the central nervous system, and thus it is unlikely that it is directly dependent upon pH of the extracellular fluid. More likely it may be the result of intracellular changes which have resulted after extracellular pH has been drastically altered.

The presence of feedback loops within the kidney itself to regulate the amount of acid or anion or cation excreted has never been demonstrated. This again would suggest that the kidney's function in this regard may be viewed as merely an exaggerated capacity of somatic cells to maintain their internal environment with certain consequent changes in the extracellular environment of these specialized cells.

Regulation of acid-base equilibrium: An important system with respect to the regulation of acid-base equilibrium of the extracellular fluid consists in the changes in concentration of carbon dioxide, carbonic acid and bicarbonate as a result of respiratory activity. As was mentioned earlier bicarbonate is the principal buffer in the serum. Although in a static system at the normal pH of body fluids the carbonic acid-bicarbonate buffer ratio is not optimal for removal of acid from the extracellular fluid, nevertheless

the efficiency is much improved by the fact that large amounts of carbon dioxide can be effectively excreted by increasing respiratory activity. This is accomplished by means of a feedback control system which is composed of sensory receptors located in the carotid and aortic bodies adjacent to the blood stream and in the respiratory center in the diencephalon. These receptors are sensitive to both changes in pH and in $p\text{CO}_2$ in their external environments. It is not known whether changes in pH (which is a weaker stimulus than $p\text{CO}_2$) may be the primary stimulus but that changes in $p\text{CO}_2$ will result in a change of pH and thus stimulate the receptors. An increase in $p\text{CO}_2$ will cause an increased rate of respiration thereby tending to lower the plasma $p\text{CO}_2$ level and shifting the ratio of bicarbonate to carbonic acid, thus improving the efficiency of the buffer. The time constant of this system is quite short, being in the order of minutes. However, the system cannot be considered to be linear, because the mode of operation will change upon exposure to prolonged stimulation by elevated $p\text{CO}_2$ levels. Under these circumstances the sensitivity of the receptors or the gain of the amplifier will diminish. (This, however, may simply be the result of the readjustment of the carbonic acid bicarbonate concentration ratio.)

Simulation of the System of Electrolyte Regulation: Design Philosophy of the System

One of the heretofore insurmountable barriers against attempts to visualize the design philosophy of renal mechanisms was the gigantic difficulty of experimental verification of any hypothesis by experimental measurement. The large number of non-linear variables which participated in every situation investigated could not be evaluated by human computation alone. The advent of computers has brought this aspect of the problem under realizable control.

The hypothesis regarding the design philosophy of the renal regulation of electrolytes follows directly from the considerations outlined in the previous section.

Since the excretion of hydrogen ions appears to be under no esoteric regulating influence with an obscure design philosophy, then perhaps the pattern of electrolyte excretion is determined solely by equilibrium and free energy considerations. Thus, if one took into account (1) the concentration gradients between plasma and tubular fluid of all pertinent ionic species --probably no more than **fifty**, (2) the addition of free energy to the system in a predictable, orderly fashion as supplied by the pumping force of the heart and the active transport of sodium and (3) the necessity for maintaining osmotic and electrical equilibrium across the tubular cell membrane, this may constitute sufficient information to

accurately predict the pattern of electrolyte excretion of the kidney under a variety of conditions.

The computations required would, of course, be formidable; the nature of this problem is such, however, that it would probably lend itself well to mathematical programming. Although changes in concentrations across various types of gradients are non-linear functions, the change in free energy that accompanies these transfers is generally linear. By assigning to each of the functions which determine the transfers of ionic and molecular species across different gradients a free energy value which can be computed, a functional can be obtained for the free energy of the entire system, which, at equilibrium, will be a minimum.

By linear programming techniques it is feasible to determine the minimum of this functional; the information thus obtained can then be compared with the large body of data which has already been accumulated regarding the relationship of, for example, sodium, hydrogen and potassium ion excretion. If this data would agree with that which has been predicted by the minimum value of the functional, it would then be possible to simulate the system, and predict its behavior by substituting appropriate values into the functions from which the functional was derived.

This approach appears at present to be the most promising, although its realization may necessitate revision of many of our concepts of renal function, and a large amount of experimental work.

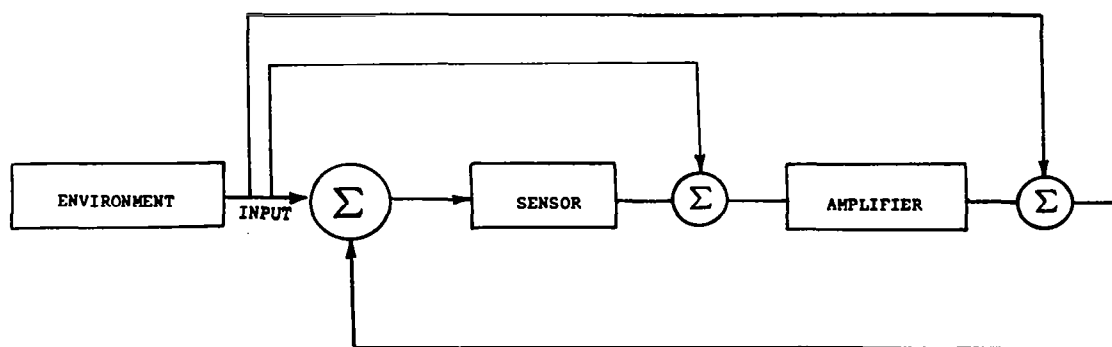


Fig. 1. Biological amplifier operating in a biological environment.

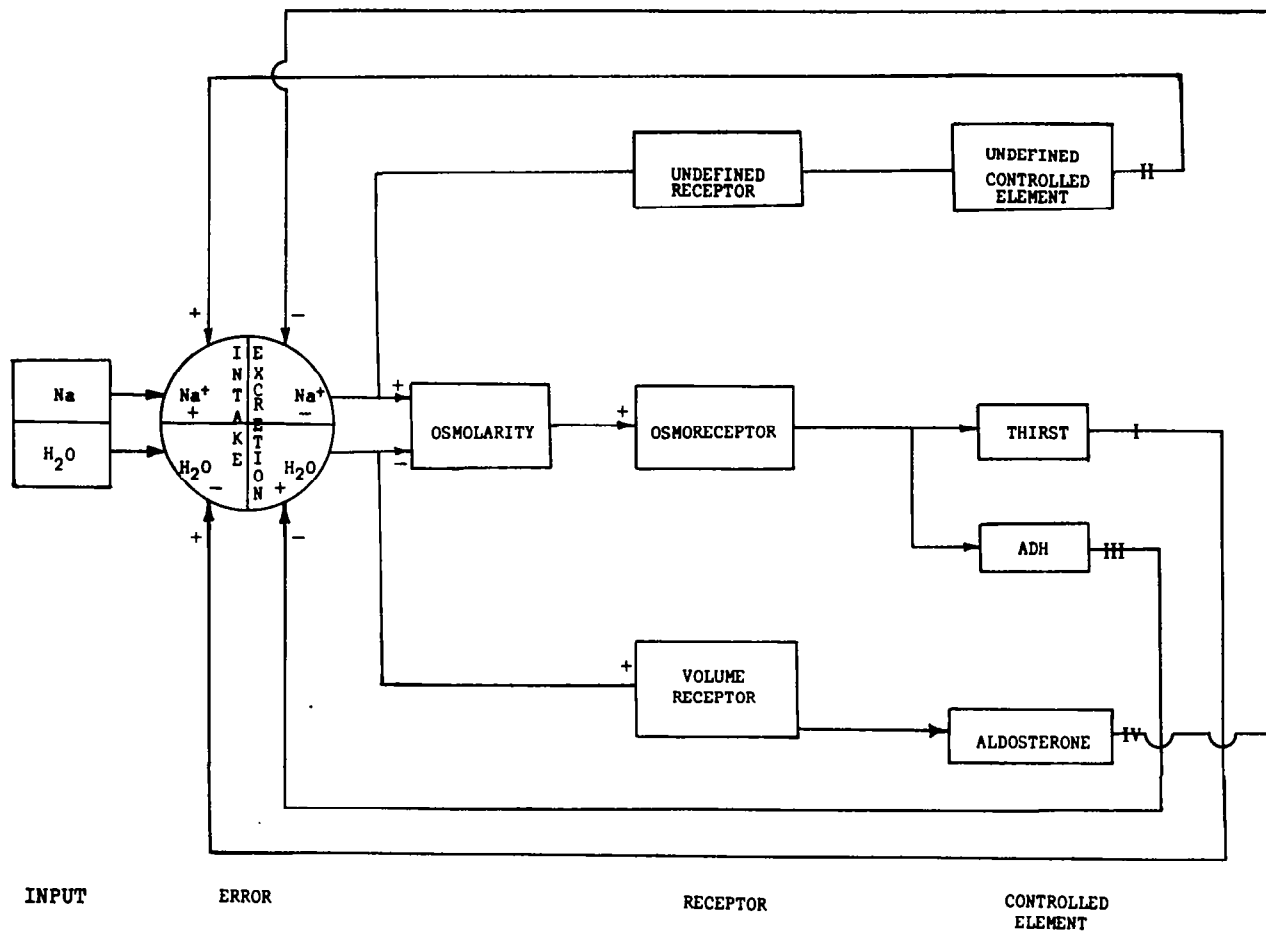


Figure 2. Control systems regulating volume of body fluids.

BIBLIOGRAPHY

1. A REVIEW OF CERTAIN BASIC ASPECTS OF THE REGULATION OF BODY FLUIDS

Body Fluids: Composition and Partitions

1. Gamble, J.L. Chemical Anatomy, Physiology and Pathology of Extracellular Fluid. Harvard Univ. Press, Cambridge, Mass., 1950.
2. Manery, J.F. Water and electrolyte metabolism. *Physiol. Revs.*, 34:334, 1954.
3. Welt, L.G. Clinical Disorders of Hydration and Acid Base Equilibrium. 2nd Ed., Little Brown and Co., Boston, Mass., 1959.
4. Elkinton, J.R. and T.S. Danowski. The Body Fluids, Basic Physiology and Practical Therapeutics. The Williams and Wilkins Co., Baltimore, Md., 1955.

Exchange of Water and Electrolytes between Various Compartments of the Body

1. Clark, W.M. Topics in physical chemistry. 2nd Ed. The Williams & Wilkins Co., Baltimore, Md., 1952. Chapter 8.
2. Welt, L.G. Water balance in health and disease. In: Diseases of Metabolism, ed. G.G. Duncan. 5th Ed., W.B. Saunders Co., Philadelphia & London, 1964. p.449.
3. Morier, J.H. and M. Stook. Edema -Mechanisms and Management (Hahnemann Symposium). W.B. Saunders Co., Philadelphia & London, 1960.
4. Ussing, H.H. Transport of ions across cellular membrane. *Physiol. Revs.* 29:127, 1949.
5. Edelman, I.S. Transport through biological membranes. *Ann. Rev. Physiol.*, 23:37, 1961.
6. Edelman, I.S. and J. Leibman. Anatomy of body water and electrolytes. *Am. J. Med.*, 27:256, 1959.

Maintainence of Acid Base Equilibrium

1. Davenport, H.W. The ABC of acid base chemistry. Edwards Brothers Inc., Ann Arbor, Mich., 1947.
2. Welt, L.G., ed. Essays in metabolism. Little Brown & Co., Boston, Mass., 1957.
3. Comroe, J.H., Jr., R.E. Forster, II, A.B. Dubois, W.A. Briscoe, and E. Carlson. The lungs. Clinical Physiology and Pulmonary Function Tests. Review Book Publishers, Inc., Chicago, Ill., 1955, Chapter 6.
4. Robinson, J.R. Metabolism of intracellular water. *Physiol. Revs.*, 40:112, 1960.
5. Darrow, D.C. and S. Hellerstein. Interpretation of certain changes in body water and electrolytes. *Physiol. Revs.*, 38:114, 1958.
6. Schmidt-Nielsen, B. Urea excretion in mammals. *Physiol. Revs.*, 38:139, 1958.
7. Thorn, N.A. Mammalian antidiuretic hormone. *Physiol. Revs.*, 38:169, 1958.
8. Robinson, S. and A.H. Robinson. Chemical composition of sweat. *Physiol. Revs.*, 34: 202, 1954.

Physiology of the Kidney

1. Smith, H.W. The Kidney, Structure and Function in Health and Disease. Oxford University Press, New York, 1951.
2. Smith H.W. Principles of Renal Physiology. Oxford University Press, New York, 1956.
3. Strauss, M.B. and L.G. Welt. Diseases of the Kidneys. Little Brown & Co., Boston, Mass., 1963.
4. Davis, J.O. The control of aldosterone secretion. *The Physiologist* 5:65, 1962.
5. Gulyassy, P.F., C. Van Ypersele Destrihou and W.B. Schwartz. On the mechanism of nitrate induced alkylsis, the possible role of

selective chloride depletion in acid base regulation. J. Clin. Inves., 41:1850, 1962.

6. Schwartz, W.B. and A.F. Relman. The kidney and potassium metabolism, in Renal Disease, ed. by D.A.K. Black. Blackwell Scientific Publications, Oxford, England, 1962, p. 553.

2. CONSIDERATIONS OF SOME ASPECTS OF CONTROL SYSTEMS OPERATING TO REGULATE THE VOLUME AND COMPOSITION OF BODY FLUIDS

1. Stark, L. and L.R. Young. Defining biological feedback control systems. Ann. N.Y. Acad. Sci., 117:426, 1964.

SECTION 5

CONTROL OF SENSATION FROM THE SKIN

by
Arthur Taub

Summary

Transduction of stimulus energy delivered to the cutaneous surface may be modified by hormonal and neural mechanisms. Conduction into the central nervous system of the information so transduced is accomplished mainly through two parallel channels, differing in their fiber size input spectra. Both channels are under local, spinal segmental, and supraspinal negative feedback control. The dorsolateral spinal cutaneous afferent system, the channel with slow peripheral fiber input is also acted upon by positive feedback from local spinal mechanisms. This channel, despite its slow peripheral fiber input, is the more rapidly conducting within the central nervous system. As slowly conducting peripheral nerve fibers (in the myelinated group) have higher thresholds, it appears that the dorsolateral spinal system may serve rapidly to signal the brain of the presence of a high-intensity cutaneous stimulus.

The implications of these observations for thalamic and cortical processes are discussed.

CONTROL OF SENSATION FROM THE SKIN

I. Introduction

A. Overview

In this report the function of the skin as receptor and encoder for mechanical and thermal stimuli will be considered. The possibilities inherent in this function for utilization of the sensory capability of the skin as a surrogate for or as an addition to other means of sensory reception will be discussed. The structure and function of cutaneous sense organs, the possibilities for modification of peripheral encoding processes through central nervous system function, the variety of channels through which peripherally encoded neural information may enter the central nervous system, and the means by which the central nervous system is thought to process this information, will also be presented from the standpoint of current neurophysiological investigation. The relationship of recent neurophysiological studies to some advances in cutaneous psychophysics will be touched upon.

In a variety of lower organisms cutaneous photo-receptors are present which appear to be related to behavior. Although human skin is responsive (by pigment change) to ultraviolet radiation, physiological evidence for the existence of a cutaneous receptor responsive to electromagnetic waves in the visual region of the electromagnetic wave spectrum has not been forthcoming. Recent reports of visual discrimination, and , in fact, color discrimination using the skin as the primary receptive organ, have not as yet been subjected to physiological analysis and are excluded from this report. Recent evidence denies the existence of such phenomena (124).

B. Redundancy Reduction

Thermal and mechanical stimuli impinging upon the skin are encoded at peripheral receptors of varying complexity and sensitivity into a series of nerve impulse trains which then, after a delay of several milliseconds, arrive at the spinal cord through the dorsal roots, when the skin of the trunk is stimulated, and at the trigeminal nerve nucleus in the pons and medulla, when the facial skin is stimulated. At this first synaptic level, and indeed at the peripheral receptor itself, a wide variety of feedback interactions (degenerative and regenerative) between the central nervous system and the encoding mechanism are demonstrable. These interactions are functions both of on-going and of past experience. The precise

nature of the nerve impulse code, however, is unknown. It had long been thought that the major function of nerve cell masses interposed between the peripheral receptor and the cerebral cortex was essentially to "relay" the information encoded at the peripheral receptor to "higher" levels. In recent years the mutability of encoding processes at the peripheral receptor at first synaptic (108) and at higher synaptic levels has led to the notion that the function of neuronal relays is to reencode information in some fashion (9), so as to provide for more effective utilization of information at central nervous system levels with the limited neural apparatus available. Several observations serve to underscore and reinforce this concept. Most significant is the demonstration that in the earliest stages in the amphibian visual data processing system, and in higher levels in the mammalian visual data processing system, neurons exist responsive only to relatively complex stimuli involving such aspects as curvature, directionality of motion, net dimming, and angle from the vertical (75). Barlow (9) has suggested that a major purpose of the encoding process at early synaptic levels may be considered to be the elaboration of an optimal code, that is, a code such that message redundancy is reduced to fit the reduced channel capacity available at higher central nervous system levels. The observation, both in the visual and the somasthetic systems, that from periphery to higher synaptic levels the average frequency of spontaneous or evoked nerve

impulses progressively decreases (9) is in keeping with this hypothesis. A major purpose of this report is to elucidate the possibilities, suggested by this hypothesis and by recent neurophysiological evidence, for modification of encoding processes at various levels of the central nervous system.

II. Cutaneous Psychophysics (105)

A. Psychophysical Principles

1. Weber-Fechner Law and Polynomial Approximation. In the early part of the nineteenth century, Weber, in experiments on lifting of weights, demonstrated that the perceptability of variations in the intensity of a stimulus depended upon the level of pre-existing stimulus intensity. That this is generally true can be seen when it is considered that the small addition of an area to a postage stamp would be much more easily discernable than the equal addition of area to a large poster. Thus, if one stipulates the variation in stimulus intensity required to produce a noticeable sensory change it is observed that

$$\Delta \phi = k \phi$$

where ϕ is the existing stimulus intensity and k is a constant. Fechner assumed that noticeable changes in sensory magnitude associated with these changes in stimulus magnitude were identical, no matter what level of existing sensory magnitude obtained. That is, he assumed that the same change in sensory magnitude occurred no matter whether the stimulus level varied from one to ten or from a hundred to a thousand. He assumed then, that $\Delta \psi$, the change in sensory magnitude

associated with a change, $\Delta\phi$, in stimulus magnitude, was a constant

$$\Delta\psi = k$$

thus, ψ , the sensory magnitude, was equal to $k \log\phi$. It thus emerged that as stimulus intensity increased geometrically sensory intensity increased arithmetically !

Now, $\Delta\psi$, the change in sensory magnitude, corresponding to $\Delta\phi$, the change in stimulus magnitude, is a measure of the variability of observer error. There was no a priori reason to state that this error would or should be constant all along the magnitude scale. And, in fact, for most physical variables it is the percent error which remains constant along the magnitude scale.

Fechner's results imply a logarithmic dependence of sensory magnitude upon stimulus intensity. Recent experiments have indicated that this dependence may better be approximated by a polynomial relationship

$$\psi = k \phi^n$$

where the exponent depends upon the particular sensory continuum being measured (Fig. 1). As an example, one may note what occurs when the luminance of a spot of light is doubled as compared

with doubling of the magnitude of a sixty cycle current passing through the fingers. Doubling the luminance of a spot of light in a dark field increases the apparent brightness only about 25% but doubling the current through the fingers makes the sensation of shock appear to be about ten times as great. The exponent n has the value 0.33 for brightness and 3.5 for shock (Fig. 2). The value of k depends on the units chosen. The power function representation of the stimulus sensory magnitude relationship has been further modified in recent years to include threshold effects so that a more general expression for the relationship between stimulus intensity and sensory magnitude is

$$\psi = k(\phi - \phi_0)^n$$

where ϕ_0 is a constant value corresponding to threshold.

2. Pitfalls in Neuropsychological Correlation. It has been possible to determine such stimulus-response relationships for a variety of stimuli to the skin. Receptor structures on the surface of the skin transduce stimuli to the cutaneous surface into a complex pattern of neural impulses. Recent years have seen the development of a neuropsychological program by some investigators in which parameters of the neural impulse pattern are related to the stimulus-response function determined for the particular kind of stimulus used to stimulate the cutaneous receptor (89). In these studies

it is implicitly assumed that the neural steady-state pattern at the receptor is a direct correlate of the sensory magnitude and that the major transformation between stimulus and sensory magnitude occurs at the stimulus-receptor interface. Thus, when a power series relationship between the stimulus intensity and an arbitrarily selected "characteristic" steady-state portion of the neural frequency response of the receptor to a cutaneous stimulus is found to obtain, and when such correspondence is found closely to parallel the polynomial function stimulus-response relationship between stimulus intensity and overall sensory magnitude, a causal relationship has been postulated to exist between the neural impulse steady-state frequency and the sensory magnitude (Fig. 3), in which the sensory magnitude is a linear function of the neural impulse steady-state frequency. In this report we shall not discuss in detail the evidence relating to the truth or falsity of this latter hypothesis. However, it is important to point out that it is unlikely that the nervous system in fact rejects all transient aspects of a response and operates only upon the later-occurring steady-state value. It is as essential to determine the destiny and significance of those portions of the nerve response to cutaneous stimulation which do not bear a polynomial function relationship to the stimulus, as it is to analyze those portions which do bear such a relationship. Analysis of the neural event at the periphery of the central nervous system can provide only

a working hypothetical construct upon which to erect a theoretical picture overall of central nervous system function, which is reflected by polynomial approximation to a sensory input-output function. It does not of itself provide an explanation of the mechanism of such function.

3. Prothetic and Metathetic Continua (105). Recent psychophysical investigations have revealed the presence of two forms of sensory continua, the prothetic and the metathetic. When measured in subjective units, prothetic continua show essentially no difference in the just noticeable stimulus difference over the range of sensory magnitude. In metathetic continua, the just noticeable stimulus difference grows rapidly larger as the sensory magnitude is increased. Typical examples of the two continua are pitch which is metathetic and loudness which is prothetic. On the pitch continuum, a metathetic continuum, resolving power is uniform over the sensory spectrum of pitch, whereas on the loudness continuum, the resolving power is non-uniform being much smaller at higher sound intensities. Prothetic continua (loudness, brightness and subjective intensity in general) seem to be concerned with "how much". Metathetic continua (pitch, apparent azimuth, apparent inclination) have to do with "what kind?" or "where?", i.e. position. Corresponding to these two functional classes there appear to be two basic physiological mechanisms. As Stevens (105) points out,

sensory discrimination can be, in the model, considered to be mediated by either of two processes, one additive and the other substitutive. For example, an increase in loudness is detected when excitation is added to excitation already present. A change in pitch is detected when new excitation is substituted for excitation that has been removed. With respect to the skin it is possible to tell when a light pressure changes to a strong pressure at a given point, (addition of excitation) and in addition it is possible to determine when a stimulus is removed from one to another location (substitution of excitation). Thus the sensory correlate of change in pressure on one point of the skin is a prothetic continuum, whereas the sensory correlate of the change of location of a stimulus is metathetic. In considering neurophysiological phenomena relating to sensation from the skin one must keep in mind that information required both for prothetic and metathetic sensory experience is provided by the identical receptors and that the differences of sensory experience resulting from the stimulation of these receptors is accomplished by different operations upon the sensory input provided to them. In the neurophysiological portion of this report we shall discuss those factors which control on the one hand the portions of the neural response pattern relating to stimulus intensity, and on the other hand those portions of the neural response pattern relating to stimulus localization, which would correspond to prothetic and metathetic sensory continua, respectively.

4. Field Variation: Parallel between Audition and Somesthesia. Earlier work on cutaneous psychophysics demonstrated the existence of discrete sensitive "spots" acting as receptors for individual sensory "modalities" and a variety of studies demonstrating the sensitivity and responsiveness of discrete portions of the skin are well known and will not be further discussed in this report (81). We shall, ^{however,} present recent results in psychophysical research for the information they provide concerning the possible activity of nervous mechanisms controlling perception of cutaneous stimuli.

Melzack, Rose and McGinty (78) studying temperature sensation of the skin found the sensitivity to temperature varied discontinuously in the skin. The "receptive fields" for particular stimuli varied in area from measurement to measurement with time, either "fragmenting" or "coalescing" (Fig. 4). Fixed thresholds or fields were not present. In addition, a stimulus, perceived as pleasantly warm when applied to large regions of the skin, was felt as a pricking or stinging sensation when applied over smaller areas. Thus, mechanisms controlling the perception of intensive cutaneous stimuli, do not normally maintain the sensitivity of the skin at constant values.

Von Békèsy (10) demonstrated a striking resemblance between cutaneous and auditory psychophysics with respect to a wide variety of phenomena, among which were the increase

with time of the magnitude of a vibratory sensation, and the variation of threshold of vibration with frequency. Von Békèsy (10) showed also that when sensation from a point source of vibration was compared with that from a wire frame near threshold, an identical change in stimulus amplitude resulted in a larger change in sensation intensity for the point than for the frame (Fig. 5), a result in the tactile sphere analogous to data in the thermal sphere presented by Melzack, Rose and McGinty (78). The inverse of this situation, that is, where the apparent size of the perceived object changed with the intensity of stimulation, that is the larger objects seemed smaller, was also demonstrable, and was attributed by Von Békèsy (10) to the increase in "lateral inhibition" for fast-occurring changes in the stimulus. The neural mechanisms for "lateral inhibitory" phenomena will be discussed below (Fig. 6).

The phenomena of experienced directionality as a result of time delay in the application of two successive stimuli, so evident in audition, was vividly manifest in the experiments of Von Békèsy (10) for cutaneous sensation when two vibrators were applied to the arm. When the time differences between stimulus deliveries were large, a separate stimulus was felt under each vibrator, but the areas of sensation fused when the time delays were shortened. The time difference necessary to make a fused sound image travel from one side of the head to the other was identical to that required to make a sensation of vibration travel from one area on the skin to another. This

parallelism between the mechanism of processing of cutaneous and auditory data is profound and may ultimately result in the demonstration of common physiologic mechanisms underlying them.

5. "Funnelling" and Contrast. Phenomena on the skin akin to the "Mach's band" effect (law of heightened visual contrast) were also demonstrable in Von Békèsy's studies (10). For example, if a frame were placed upon the skin lightly, only the edges were perceived. If its central support were struck with a hammer then a diffuse sensation, apparently in the center of the frame, was reported. On the basis of this and other observations, Von Békèsy advanced the concept of "funneling" as a paradigm of the way in which sensory input is processed. In an experiment illustrating this concept (Fig. 5), Von Békèsy demonstrated that when the sensation magnitudes of five vibrators, linearly placed upon the arm, whose vibration frequencies were an octave apart, were equalized separately by adjusting their vibratory amplitudes, only the vibrator at the center of the linear array was perceived when all were activated simultaneously. The other vibrators were not perceived. The central vibrator, which was perceived at its correct frequency, was perceived to be vibrating with an amplitude greater than its amplitude as perceived when vibrating alone. Thus, "the inhibited sensation somehow contributed to increasing the loudness or the sensation magnitude of the central sensation". The term "funnelling" was introduced to express the fact that

although four of the vibrators did not contribute to the sensation of vibration frequency, and in fact were not perceived in themselves at all, they did contribute to the sensation of vibration amplitude of the vibrator whose frequency was perceived.

6. "Funnelling" and Amplitude. In another experiment a frame vibrating at 50 cps was placed upon the arm and its vibration intensity was adjusted to 50 db above threshold. The perceived lateral spread of the sensation induced by the frame was as long as was the frame itself. If a series of clicks, instead of a steady vibration, was introduced onto the frame, the perceived lateral spread of the click sensation was much smaller than without the clicks. Superimposing a periodic click stimulus upon a constant amplitude of vibration caused a pronounced recession of the perceived lateral spread of the vibration, which recession increased as the click amplitude was increased above threshold. The vibration was perceived as of greater amplitude than when occurring alone. Paraphrasing Von Békèsy, the clicks "funnelled" the vibration sensations into their own constricted receptive field. In addition, the energy provided by the clicks increased the vibration sensation. Other experiments indicated the persistence of the "funnelling effect" for several seconds.

Von Békèsy (10) derived a theoretically expected shape for a "funnelled" sensory field by assuming that associated with each excitatory stimulating point an inhibitory influence was

exerted upon the surrounding points (Fig. 6). It was possible to derive approximations of perceived shape and amplitude of stimulus configurations by the assumptions of appropriate geometries, and excitatory and inhibitory intensities, for individual points.

7. "Funnelling" and "Afferent Inhibition" in Physiological Processes. "Lateral", "surround" or "afferent" inhibition has been described in physiological experiments using cutaneous stimuli in a number of loci in the central nervous system: the dorsal column nuclei (94), somatosensory cortex, the ventrobasal thalamic nuclei (88), and in a dorsolateral spinal cutaneous afferent system projecting to the lateral cervical nucleus (108). This effect has heretofore been considered to play a role primarily in delineating receptive fields by heightening of contrast. It now appears from the work of Von Békèsy (10) that it may be of significance for control of the sensory amplitude as well. The presence of a mechanism for modifying "funneling" could have as its purpose, not only a probable increase in the accuracy of location of stimulation (a metathetic effect) but, possibly, improvement in the discriminability of a stimulus from ambient "random" activity of relatively high amplitude (a prothetic effect).

The methods by which the central nervous system may produce lateral inhibition and funneling and may modify them, will be discussed below.

B. Applied Psychophysics

1. Temporal "Funnelling". Suppression of one cutaneous input by another is manifest not only in a spatial, but also in a temporal sense and over periods of time much longer than known refractory periods of neurons. The introduction of a vibratory repetitive input produces bursting (that is, "excitation" followed by "inhibition") in second-order central cells in the dorsal horn of the spinal cord (117), in contrast to the prolonged after-discharge in these cells produced when a single pressure pulse is applied to the skin. Such relatively prolonged inhibitory effects probably underly the rise in threshold to a tactile stimulus observed when the stimulus is delivered within the circumference of a vibrating annulus (79, 117) (Fig. 7) and provides a physiological explanation for the observation that a perceived frequency of cutaneous vibration drops 3 octaves for 60 db increase in vibration intensity (10). Rosner (102) has shown that although two stimuli to the identical cutaneous locus cannot be resolved when they are separated by less than 15 msec, vibration up to 100/sec rather than 70/sec, the expected limit, is detected. Rosner (102) did not study the frequency of the perceived sensation at the upper limit. It thus may be that although the perceived frequency, may, in fact, be limited, a "temporal" funnelling effect produces a sensation perceived as at a lower frequency.

2. Cutaneous Speech. Considerations of frequency limitation also perhaps explain why the effort to obtain "cutaneous perception of speech", explored extensively in the recent past, was not a notably successful enterprise. "Cutaneous perception of speech", in the sense of voice made to impinge upon the skin by loudspeaker, has generally been unsuccessful and so has encoding by a multiplicity of vibrators placed on the skin. The effort to determine the limits of discrimination of loci to which the integumentary system and its neural apparatus can be pushed has been mainly the work of a group at the University of Virginia under the direction of Geldard (35). They have been able to show that up to six separate cutaneous vibratory loci can be discriminated simultaneously. These investigators have been able to define the correlated of the cutaneous analogue of visual "phi" (the illusion of perceived motion, as in cinematography), and so to produce a bizarre sensation of motion on the skin, such that the observer is perceived to be at its "center". This latter effect may occur even when stimuli overlap. It is critically dependent upon the sequence of stimulus presentation; and may also involve "funnelling" in a way as yet unknown.

C. Summary Psychophysics and Neurophysiology

The previous discussion of recently observed psychophysical phenomena provide the experimental observations which a consistent theory of neurophysiological mechanisms must

explain. At the same time, experimentation directed specifically to the explanation of these effects has not been forthcoming. It is quite likely that only following a parallel development of a purely "neurophysiological" framework, that is, the development of a framework of theory and experiment conducted almost without reference to the perceiving animal, will the two fields merge. In this spirit, we continue our description and explication of neurophysiological phenomena in the following chapters.

III. The Peripheral Receptor

A. Structure and Function

1. Receptor Subdivisions. Recent years have seen increasing information developed concerning the structure and function of cutaneous peripheral receptors (121). The skin (Fig. 8) contains two major groups of receptors, one organized into characteristic anatomic configurations, the "focal" receptors (hair, various corpuscles, "disks", etc.) which generally originate from large myelinated fibers, and the other unorganized and "diffusely" distributed "bare nerve endings" which arise from both myelinated and unmyelinated axonal stems. Individual "focal" receptors are usually responsive to one or to two specific forms of stimulus energy (such as, e.g., touch and/or heat exchange) although their ultimate responses may be influenced by such factors as the ambient temperature and the blood supply to the tissues. Individual "diffuse" receptors have been shown to respond to several of a wide variety of stimulation energies (hair, tactile, thermal and various combinations) although restriction of response to one form of stimulus energy change is present in some cases.*

* A note about an "abuse of language": A receptor, fiber, or system is said to "respond" to a stimulus energy if the frequency of nerve impulses produced by it increases with increase in stimulus energy.

2. "Focal" Receptors. There are three varieties of mammalian skin: hairy, glabrous (e.g. palms) and membranous (e.g. lips). With the exception of the cornea, which is innervated by a "diffuse" network only, all skin is innervated by both "focal" and "diffuse" receptors (119, 120). "Focal" receptors are present to the greatest extent in those skin types where epithelial ridges (so called "rete pegs") are developed in the substructure.

The development of focal receptors is a recent one in vertebrate evolution. The innervation of the skin of *Amphioxus*, a primitive chordate, is generally similar to that found in the mammalian cornea, a "diffuse" network of bare nerve endings. Innervated "tactile spots", "focal" receptors, are found to be associated with reptilian scales, and it is thought that the various focal receptors in mammals, such as the "tactile disks" and the "mammalian end organ" are homologous to these structures, and thus represent a direct evolutionary line.

Parallel to the development of the peripheral receptor in evolution is an associated development of central data processing systems to deal with the information presented by the receptor. It has been suggested that the mammalian hair follicle sensory system probably evolved synchronously with the development of the dorsal column nuclei, in which may be recorded responses to hair movement. These latter structures, located at the junction of the medulla and spinal

cord of higher vertebrates, are of the greatest significance in the processing of cutaneous information as we shall see. They have evolved in synchrony with their receptor apparatus (Fig. 9).

a. Hair follicle. The major "focal" receptive end organ of the skin is the hair follicle in which are located nerve endings sensitive to movement of hairs. The sensitivity of the sensory receptors associated with the hair follicle must be maintained in the face of a continuous cycle of hair follicular growth and variations in tissue turgor associated with the hair-growth cycle (74,106). The hair follicle produces its hair cyclicly, with a short resting phase following upon a longer active phase (95). In man the active growth phase may last from 6-8 years, and up to 20% of scalp hairs are in the resting phase at any one time. During the growth cycle the hair increases in length and thickness, thus varying the mechanical properties of the transducer by changes in mass and lever-arm. Despite these changes sensation from hair stimulation is in general unchanged.

In some forms (rabbit, rat, mouse) the hair growth cycle is synchronous over wide portions of the skin, and decreased skin turgor is present during the resting phase (74,106), thus decreasing elastic resistance to movement, and varying overall end-organ sensitivity. Vibrissal (74) and tylotrichal (74) (specialized forms) hair are distinguished by the presence of

a complex venous or capillary sinus, which is said to be capable of variation in filling under neural control during different phases of the hair growth cycle (74). In this manner it acts as a mechanism to compensate for elastic and damping changes due to varying tension of the hair shaft against the surrounding follicular nerve network during the growth cycle.

The neural network innervating the hair follicle (Fig. 10) and serving as its sensitive receptor is arrayed in two layers, an inner longitudinal portion and an outer circular portion, both of which represent unmyelinated terminals of myelinated axons. These axons gradually increase in size as they approach the myelinated portion of the stem axon which will carry the transduced information to the central nervous system. The hair follicle possesses another sense organ in addition to the neural apparatus surrounding the hair shaft. This organ, a mound of innervated epidermis near the follicular orifice, is provided with a specialized erectile tissue also perhaps capable of engorgement and equalization of receptor sensitivity against variations in tissue turgor. It is termed the "tactile disk of Pincus" and, in the older literature, the "Haarscheibe". It has been claimed that the follicular plexus and the tactile disk are innervated by a single large axon, that is, that they are in fact one receptor (74). The function of the tactile disk will be discussed separately below.

Quantitative information concerning the hair follicle and its innervation has been provided by the studies of Weddell et al. using the skin of the rabbit ear and the cat foreleg as a model (119,120).

Some data: On the surface of the rabbit ear one hair follicle may give rise to up to 7 hairs. 10^5 hairs are present on the surface of the ear, in a complex nonperiodic pattern, which is nonetheless similar from animal to animal. Hair shafts emerge from 4×10^4 orifices, the latter grouped into 1.3×10^4 groups. The average skin area occupied by a group of orifices is 0.25 mm^2 . The average distance between hair follicles is 0.35 mm. No hair follicle is innervated by less than two dorsal root axons, and some follicles are innervated by 20-30 such axons. On the average each dorsal root axon innervates 80 hairs, some innervating more than 100 hairs. These stem axons are myelinated and on the rabbit ear generally are larger than 6 microns in diameter. On the cat foreleg, stem axons to hair follicles may be as small as 4 microns in diameter. The total number of myelinated stem axons supplying the rabbit ear is about 0.6×10^4 .

The grouping of several hairs into one hair follicle makes it extremely unlikely that a single neural apparatus

surrounding an individual hair shaft is ever stimulated, under natural conditions. Also the proximity of hair follicles is such that under natural conditions it is also unlikely that a single hair follicle system is stimulated. Further it is clear that with natural stimulation, i.e. manipulation of the hair, it is almost impossible to stimulate a single dorsal root axon. Finally, analysis of information carried along a dorsal root fiber innervating a hair follicle must take into consideration that up to 100 hair follicles are simultaneously providing information concerning their status to the same stem axon.

What this implies is that the pattern of nerve impulses in an individual dorsal root axon, that is, the pattern of firing in the only channel from the periphery into the central nervous system, must, under normal conditions be extremely complex, even if, (and this itself is not yet proven completely) the angle of displacement of an individual hair shaft in its follicle were transduced into pulse frequency. This comes about primarily because of the interaction of, as we have said, up to 100 inputs upon the branch points leading into the single axon. Branch points are unusually well suited to act as modulators of neural patterns, because as a nerve impulse enters a branch, it may draw current from a much larger region of membrane than it would have had it been propagating down an unbranched path. This current, however, may not be of strength sufficient to ensure

propagation down the branches. The slightest degree of refractoriness in the vicinity of a branch, as a result of the passage of a previous impulse, adds to this possibility for control. It may be seen from this simple example alone, that whatever the coding mechanism at the transducer, a further encoding must occur by virtue of the dorsal root branching into which the receptor is discharging its impulses.

The hair follicle receptor may send its output into the nervous system via one of two groups of fibers, conducting at either c. 25 meters/sec or 100 meters/sec (51,52). It is silent when enstimulated, and is generally unresponsive to chemical stimuli (31) (except for the vibrissae of the cat which are responsive to high concentrations of potassium) (30). Adaptation (that is, decrease in response with maintained stimulus intensity) is rapid, and the receptor is resistant to fatigue (51,52) (that is, decrease in response with repeated activation).

b. Pacinian corpuscle (53). The Pacinian corpuscle has been shown to be capable of transducing vibration from the skin. This end organ, consisting of a nerve ending embedded in an onion-skin shaped capsule is surprisingly specific in its response to displacement which specificity the nerve ending retains when the encapsulating connective tissue is removed (67). It is insensitive to chemical stimuli.

When recording from single units in a dorsal root connected peripherally to a Pacinian corpuscle, it was found that these end organs are sensitive to movement of the skin over the entire extremity and thus are probably responsible for reports of single units with extremely wide receptive fields on the skin. In addition it was found that the Pacinian corpuscles were highly sensitive to vibrations up to 400 cycles per second, which was not found with other cutaneous mechanically sensitive end organs (53).

c. Mammalian end-organ; Meissner corpuscle

In special sites in hairy skin, but chiefly in the glabrous (bald) skin of the paw of the cat and other lower mammals (e.g., dog, sheep, rabbit, guinea pig), a differentiated end-organ exists, termed the "mucocutaneous" or "mammalian end-organ"⁽¹²²⁾. This is a multiply convoluted concentric capsule, with an expansion of myelinated nerve, resembling an embryonic Pacinian corpuscle. It is found in the dermal papilla, below the epidermis. Organs of this type are easily seen at the base of each digit along the flexor tendon, where true Pacinian corpuscles are also present. This "end-organ" is found to be similar in a wide range of feline species. A large axial nerve fiber courses the entire length of the end-organ. About this is a homogeneous zone, and peripherally, is a connective tissue capsule composed of several layers of concentric fibrous tissue resembling the true Pacinian corpuscle. The end-organ is located in mucocutaneous tissue,

and is often associated with "tactile disks". Primates, including man, do not possess this end organ, the chief mucocutaneous end-organ in these forms being the Meissner corpuscle, a complexly convoluted neural tuft, with a capsule. It was maintained at one time that the Meissner corpuscle in the primate hand was the result of degeneration of bare nerve endings due to continual trauma (see below on the subject of neural turnover). That this is not necessarily the case is indicated by the fact that the raccoon possesses no Meissner corpuscles in the glabrous skin of its prehensile paw. Instead, the mammalian end-organ is present (121,122). Thus, despite trauma, no Meissner corpuscles appear.

The function of the Meissner corpuscle is unknown.

d. Merkel disk; "end bulbs". In the skin of the lips, some of the most superficial nerve endings related to the cells of the germinal layers of the epidermis are distributed over not more than 1 mm^2 of skin surface/axon; they are derived from large axons and are surmounted by distinct swellings termed "Merkel's disks". One stem axon may innervate up to five Merkel's disks. "End bulbs" may be also seen in the exposed mucous membrane of the lips and consist of simple clubbed endings possessing Schwann-cell capsules and may be arranged in complex arborizations. The varied anatomical appearance of "end bulbs" has been explained to be a result of "neural turnover", as discussed below.

The function of Merkel's disks is unknown, but may be related to the "touch spot" described below.

e. Tactile disk ("touch spot". A characteristic epidermal elevation, appearing slightly vesicular (blister-like) in angulated indirect light, extending about 100 microns from the surface of the hairy skin, is, in fact, a complex innervated sensory structure (74,106). Frankenhaeuser (32), in the rabbit, described a pressure sensitive focal receptor which adapted slowly to tactile stimuli, and which was arrayed in multiple focal spots innervated by the same axon. Similar observations were reported by Hunt and McIntyre (51,52) in the cat. These investigators did not identify their "touch spots" anatomically. Physiologically, the sense organ was found to consist of an organized unit of 1-5 "spots" innervated by a single large myelinated sensory axon and was responsive to touch and to decrease in temperature (that is, a slow decrease in response with maintained stimulus intensity), with slow adaptation and marked fatiguability (decrease in response with repeated stimulation). Iggo (54) and his co-workers and later Tapper (107) studied this organ anatomically and physiologically, confirming previous studies, and Fjällbrandt and Iggo (31) demonstrated a striking pharmacological responsiveness of the "touch spots" to intra-arterial acetylcholine, serotonin, epinephrine and norepinephrine. The "touch spot" fires "spontaneously".

Recently Mountcastle and his co-workers have studied the touch spot in detail, and have determined its characteristic response to precisely controlled deformation of the skin. The touch spot is sensitive to deformation of less than 50 microns. A brief rise, in frequency is followed by an early steady-state and the steady-state frequency of response is nonlinearly related to the displacement. Quantitative reports of these investigations have not as yet been published. Mann and Straile (74) have identified the touch spot with the pad associated with the tylotrich follicle. Its chemical responsiveness may be related to vaso-constriction.

f. "Neural turnover" of focal receptors. "End bulbs", Meissner corpuscles, and Merkel disks have been considered by some anatomists to represent attempts at regeneration of neural structures following trauma, in a manner similar to the so-called "sterile" end bulbs of regenerating nerves described in the conjunctiva by Cajal (119,120). In man, in individuals in whom end bulbs are numerous in the skin, silver stains of fibers reveal 1 in 10-20 to be undergoing signs of degeneration or of regeneration. In children the number of fibers so affected is of the order of one in 1000, and in adults, one in 100-500. This "neural turnover" has been demonstrated not only in "focal" systems, but in the "diffuse" neural networks of the cornea and of the human carotid sinus as well. With "neural turnover" in operation, one must suppose that compensatory mechanisms are continually in effect

to maintain constancy of the characteristic response of sensory analyzing mechanisms in the face of continually varying sensitivity and number of sensory end organs. This may be of the nature of an adaptive process associated with a "noise filter" rejecting information arriving in patterns suggestive of regeneration processes occurring in the ending. One simple way of accomplishing this end would be the fatiguing of the synaptic mechanism upon which fibers connected to regenerating end organs impinged as a result of the continued unpatterned and, presumably, heightened, neural activity arising from the mechanical stimulation of the ending incidental to the regeneration process.

3. "Diffuse" Receptors (119,120).

a. Anatomy. "Diffuse" receptors are present in all varieties of skin and possess wide receptive fields. In the feline cornea the area over which the arborization of one of the stem axons innervating the cornea spreads is on the average not less than a quadrant of the cornea and adjacent bulbar conjunctiva. Terminal filaments are found at all levels of the epithelium and substantia propria of the cornea, but are most numerous just below the level of the basal layer of the corneal epithelium. In the skin proper, in the dermis and epidermis, fine networks take the form of basket-woven arrangements of varying complexity.

Although diffuse receptors may take their origin from both myelinated and unmyelinated fibers, the vast majority of their nerve endings arise from fibers which are unmyelinated until their entry into the spinal cord.

The unmyelinated system of fibers connected to peripheral receptors is characteristically different from the myelinated fiber group (34). Unmyelinated fibers range in size from 1.3 to 0.3 microns. In contrast to the myelinated fiber group individual fibers are contained in a single Schwann cell sheath. Fibers continually realign their position within the sheath in a random manner. Despite this contiguity of fibers interaction between individual fibers is minimized, as discussed by Gasser (34). The action potentials of unmyelinated fibers are twice as long as those of myelinated fibers (about 2 milliseconds) and their refractory periods are three to six times as long (3 to 6 milliseconds). The maximal response frequency of unmyelinated fibers is slower by a factor of ten (that is, 10 impulses/sec) than the maximal response frequency in the myelinated group. The significance of this fact lies in the observation that, although the frequency of firing of an unmyelinated fiber is therefore limited, for the same (realizable) firing frequency, an unmyelinated fiber may produce a more potent effect at its central terminals due to its longer action potential. In the myelinated group a reasonably general inverse relationship has been demonstrated between threshold to electrical stimulation and fiber size or conduction velocity. In the unmyelinated group

no such consistent relationship has as yet been demonstrated. Furthermore, although, in general, the electrical threshold for the unmyelinated group is twenty times that for the largest fibers in the myelinated group, there exist small myelinated fibers, the δ -1 group, whose thresholds to electrical stimulation are greater than those of most C fibers (34).

b. Unmyelinated fibers and pain. Investigation into the functional role of unmyelinated fibers is one of the most rapidly progressing fields of neurophysiological research (34). At first, unmyelinated fiber systems were thought to be responsible for the results of experiments in which struggling, vasomotor and respiratory responses to electrical stimulation were eliminated following "differential dorsal root section" in the cat (98). This procedure was thought to cut unmyelinated fibers alone (97). It was thought that unmyelinated fibers represented for the most part a "pain pathway". This concept was extended (incorrectly) in later years to imply that the only function of the unmyelinated fibers in the cutaneous afferents was to mediate pain, a concept which apparently received confirmation from several lines of evidence. First, Erlanger and Gasser (27) demonstrated that the slowest fibers in peripheral nerves, those which produce the C, or slowest, elevation in the compound nerve action potential, were unmyelinated. Then, Clark, Hughes and Gasser (17) showed that inclusion of C fibers in a peripheral nerve volley elicited

by electrical stimulation produced a marked increase in respiration and elevation of blood pressure, which was not produced by electrical stimulation which elicited a response in larger myelinated fibers alone. It was thus inferred that C fibers were most directly related to the behavioral manifestations of pain. Also, complementary evidence that the largest myelinated group of fibers was not related to pain was thought to have been at hand through the efforts of Adrian, Cattell and Hoagland (1), who produced in the frog high frequencies of firing in myelinated fibers with no escape reflex phenomena, and by the work of Gernandt and Zotterman (36) who obtained similar results in work on the mesenteric plexus in the cat. A theoretical construct now emerged in which the larger myelinated fibers were thought to provide the information requisite for the experience of touch and temperature, and the unmyelinated fibers were thought to be responsive solely to high intensities of stimulation and thus to be associated with "pain".

Recent investigation has tended to deny this concept, and to show, first of all, that no experiment has as yet in fact been performed in which unmyelinated fibers alone have been cut. The experiments of Ranson and Billingsley (98) have been plausibly shown to depend upon the destruction of the circulation to the dorsal horn of the spinal cord, seriously interfering with all nervous transmission. Furthermore, anatomical studies (25) have revealed that the peculiar segregation

of unmyelinated fibers in the lateral zone of the dorsal root, described by Ranson (97) is, in fact, not present, and the unmyelinated fibers surround the myelinated fibers in the dorsal root. Thus, in fact, the unmyelinated fibers were not "differentially cut" in the Ranson and Billingsley (98) experiment. The result in which the cutting of unmyelinated fibers led to analgesia is placed in doubt. To consider the remaining results: Experiments with electrical stimulation in which C fibers are involved always imply that C fibers deal with high threshold phenomena. This is no doubt due to the fact that, because of their narrow diameter, C fibers can only be stimulated electrically by high current densities, and thus, high voltages. Natural stimuli to the sensitive receptors, on the other hand, are responded to at very low threshold. In fact a light natural stimulus stimulates something like 95% of unmyelinated fibers, for in fact, less than 5% of unmyelinated fibers have been found to possess high thresholds when stimulated at the sensitive receptor with natural stimulus energies. Despite this, no pain response is produced with a light stimulus. Thus, one cannot say that "under natural conditions pain results when the unmyelinated fiber group is stimulated." Thus the unmyelinated group is not sufficient for pain perception. That the unmyelinated group is not necessary for pain perception is seen when it is realized that the lingual nerve, containing fibers from the teeth, contains no unmyelinated fibers, and toothache is certainly a well-known "pain" phenomenon.

Although, as we have seen, the unmyelinated group of fibers is neither necessary nor sufficient for the release of the behavioral response related to pain, the C fiber group is related in some, as yet obscure, way to "pain" effects. When, in unanesthetized man, (18) a peripheral nerve is electrically stimulated, pain is not reported until the smaller myelinated and unmyelinated fiber groups are included in the volley produced by increasing the intensity of the electrical stimulus. When the responses of the larger myelinated fibers are blocked by the application of cold to the peripheral nerve, pain is felt if only the small myelinated and C fibers are stimulated. This despite the fact that most of these fibers are connected to hair and others are connected to skin receptors with low thresholds. This all implies that some parameter of the stimulation of these fibers is in fact related to the release of "pain" processes. Perhaps it is the frequency of firing of the individual units that is responsible? One recalls now the experiment of Adrian, Cattell, and Hoagland (1), who could not get a frog to show reflex escape phenomena when they stimulated a myelinated fiber rapidly at low threshold with an air jet to get firing frequencies which they could only achieve by high intensity stimulation. They concluded that it was not the frequency of firing, but most likely the fibre type that is the determinant of the pain release phenomenon. What appears to be actually the case, however, is that both processes are involved. On the one hand, high frequencies of firing in large

myelinated fibers of itself does not release pain processes. Low frequencies of firing in unmyelinated fibers do not release pain processes. But, high frequencies of firing in unmyelinated fibers do release such effects, whether the stimulation of unmyelinated fibers is accompanied by stimulation of myelinated fibers or not. The mechanisms by which unmyelinated fibers, when stimulated at high frequency, may have the effect of releasing pain processes which they do not possess when stimulated at low frequencies, will be discussed when spinal cord processes are considered.

c. Spectrum of response of unmyelinated receptors. Investigation during the past decade has demonstrated a full spectrum of responsiveness for the C fiber group. The techniques utilized include: single fiber bundle recording, a direct method (55), and the technique of occlusion of C fiber input by antidromic electrical stimulation of peripheral nerves (23). In the C fiber group associated with cutaneous innervation in the mammalian skin, seven major response-types have been described:

- (1) Chemosensitively specific (histamine, acetylcholine, mixture of histamine and acetylcholine),
- (2) Both chemosensitively and mechanically specific,
- (3) Low intensity mechanosensitively specific,
- (4) High intensity mechanosensitively specific,
- (5) Thermosensitively specific in four subtypes, --(a) to profound rapid cooling, (b) to mild rapid cooling, (c) to mild rapid heating, (d) to profound rapid heating,

- (6) Both thermosensitively and mechanically specific in two subtypes, --(a) to mild cooling, (b) to mild warming,
- (7) Mechanosensitively specific responses in four subtypes, --(a) wide receptive field ($50 \times 50 \text{ mm}^2$), (b) small receptive field ($2 \times 2 \text{ mm}^2$), (c) hair stimulation, (d) touch stimulation. Mechanosensitive specific units all displayed a complete range of thresholds in a continuous fashion.

Thus, the unmyelinated group seems to possess a baffling array of "specificities" when compared with the myelinated group.

4. The Partitioning of the Fiber Size Spectrum in Cutaneous Reflex and Sensory Behavior. Among other differences between the unmyelinated and the myelinated group may be enumerated the different functions of the fiber size groups in both sensory and reflex processes.

A-alpha, A-delta and C fibers participate in the elicitation of the galvanic skin reflex, the A-alpha fiber range alone being sufficient to elicit a reflex (66). Evans has shown, on the other hand, the insufficiency of the A-alpha group alone for the elicitation of the pupillodilator reflex following cutaneous stimulation, while inclusion of the A-delta and C components in the peripheral afferent volley produced the reflex (28,29). These considerations point to partitioning of fiber size projections among nuclear groups responsible for different reflex activities.

5. The Partitioning of Fiber Sizes in Response to Close-Arterial Injection of Chemicals. Unmyelinated fibers are well known to be responsive to chemical stimuli (23). In the

myelinated fiber group, however, a difference in chemical responsiveness exists between those fibers connected to hair receptors and fibers connected to tactile disk receptors.

Fjällbrandt and Iggo (31) recorded from myelinated and unmyelinated single units in the feline saphenous nerve after injections of histamine (66mg), 5-hydroxytryptamine (1-20 mg), acetylcholine (20-60 mg), of mixtures of acetylcholine and histamine, and of an impure preparation of bradykinin. Some C fibers responded to chemical stimuli. Of the myelinated group, those fibers responding as "hair"units were unaffected in that their response to mechanical stimulation was unchanged. They also, displayed no "spontaneous activity". Those units, however, termed (31) "slowly-adapting pressure receptors", and which we may now identify with the anatomically characteristic "tactile spots" or "tylotrich pads" (74), responded strikingly to the chemical stimuli. Following injection of histamine, acetylcholine, and mixtures of histamine and acetylcholine, there was an initial enhancement, lasting 2-4 minutes, of both the resting discharge and the response to pressure, followed by a depression of sensitivity, with recovery after 10-60 minutes. Bradykinin enhanced the steady resting discharges in these units. The technique used by Fjällbrandt and Iggo (31) did not excite C fibers to fire above a rate of 5 impulses per second, as distinct from results in a different experimental setup of Douglas and Ritchie (23), in which closer intra-arterial injections were used, and where C fibers were stimulated to fire more rapidly.

Differences in chemical responsiveness exists not only between myelinated "hair receptors" and "tactile disk" receptors but between different groups of hair receptors as well. Vibrissae have been shown to be specifically responsive to potassium concentration, whereas the general skin hair follicles are unresponsive to such stimuli (30).

B. Information Processing in Peripheral Receptors and Its Control.

1. Sequence of Receptor Events. From a wide series of studies in vertebrate and invertebrate forms a picture of the mechanism of receptor function has emerged in which the following sequence of events is thought to occur (45):

a. Mechanical deformation or other stimulus energy, (receptors are generally responsive to at least two forms) produces an increased ionic permeability, particularly to sodium, of a specialized terminal portion of a receptor axon. This terminal is generally embedded in some form of connective tissue structural matrix which contributes resonance peaks and damping factors to the transfer relations between the stimulus input and the specialized membrane response. These may take the form of absorption peaks, elastic recoils, viscous and non-viscous damping, and standing waves. The specialized terminal structure is thought to be unresponsive to electrical stimulation, and to be unable to conduct a propagated impulse, but to be sensitive to the action of cholinergic, adrenergic, histaminergic and other, yet unknown, agencies.

b. Ionic flow along concentration gradients resulting from the permeability change produces the generator, or receptor potential. This potential generally has been studied as a response to a step or ramp function input stimulus. Two characteristic types of generator potential response have been described as a result of a step input; one in which a steady-state constant response is reached following a short overshoot, (termed "tonic"); and one in which, following a shallow rise, an exponential decline with a time constant 4-5 times that of the rising phase is observed, (termed "phasic"). It is not clear whether these two responses are derived from properties of the transducing membrane, or primarily as a result of the mechanical coupling between the connective tissue matrix or the transducing membrane, although at the present time the latter appears the more likely.

The receptor potential is increased in magnitude over a narrow range by repetition of stimuli.

Responses to ramp inputs have not been as extensively studied, although there is limited information that some receptors do in fact come close to differentiating the input, that is, maintaining a constant generator potential proportional to the rate of rise of stimulus intensity.

c. It is generally thought to be the case that conducted neural activity begins at the first or second node of Ranvier central to the transducing membrane. In a variety of

receptors a constant frequency of firing of the sensitive node is produced by a maintained generator potential through mechanisms as yet little understood.

d. The static response of the frequency of firing of axons connected with receptor elements has been demonstrated in a variety of situations to be logarithmic with intensity, although a variety of polynomial approximations have been produced which are asserted to fit the data more closely. The logarithmic approximation was thought to be of interest in that it appeared to parallel the Weber-Fechner law in which the conscious sensory response to a stimulus was thought to be logarithmically related to its intensity.

The dynamic response of neural frequency to stimulus intensity on the skin has not been studied to any significant extent. Various non-linearities have been observed, however, for example, the frequency response curves of some thermal receptors show local temperature response maxima (24).

2. Control of Receptor Transfer Function. Peripheral receptors possess a variety of functional characteristics, some of which may be reasonably considered subject to a central control function, and others to be intrinsic properties of the transducing and impulse conduction mechanism. Combined, these form the encoding mechanism for the stimulus.

a. Intrinsic receptor properties.

(1) Chemical specificity. As pointed out, vibrissal receptors display a marked responsiveness to increased blood potassium concentration (30), general skin hair receptors associated with myelinated fibers appear to be unresponsive to adrenergic, cholinergic and histaminergic substances (31), while tactile disk receptors associated with myelinated fibers show increased sensitivity and spontaneous activity following intra-arterial application of such agents, as do a small number of C fibers (31).

(2) Stimulus energy specificity. Fibers appear to be sensitive to at least two forms of stimulation energy, but not equally, so that, for example, destructive temperature changes are required to produce responses in some receptors which can easily be stimulated with small cutaneous deformations. Those focally organized receptors studied appear to be most specific in this respect, while the diffuse networks appear less so. However, even in the C fiber group receptor specificity is present in a large number of units. The notion that all "sensory modalities", as introspectively and arbitrarily isolated from experience, are associated with receptors responding specifically to one form of stimulus energy, or conversely, the notion that for no "sensory modality" does there exist a receptor transducing specifically an appropriately matched form of stimulation energy, are both untenable. What is most likely is that the totality of experience is provided

by receptors encoding specific stimulus energies, supplemented by patterns of stimuli in relatively unspecific networks. Above all, it must be noted that anatomically "unspecific" or, rather, bare nerve ending receptors, may respond in a completely "specific" manner to stimulation energy.

(3) Disposition of sensitive terminals of a neural unit in three dimensions. The topology determines stimulus encoding in a variety of ways:

- (a) The relative proportion of terminals at the periphery and center of the receptive field of the unit determine the average sensitivity (that is the number of impulses for amount of stimulus energy) of the unit to cutaneous stimulation at periphery and center.
- (b) The relative thickness, and length of individual branches, and the refractory period in each branch determines the tendency of stimuli from one or the other portion of the receptive field to block stimuli from other portions, and also determine the average frequency of impulses finally travelling along the stem axon.
- (c) In addition to blockage at branch points, antidromic firing along branch points to the periphery will produce refractoriness at sites where generator potentials activate conducted impulses, and so effect encoding.

(4) Axon diameter-threshold ratio. A general inverse relationship exists between stem axon diameter and peripheral receptor threshold for the myelinated group. In non-myelinated series, no information concerning this relationship is available as yet.

(5) Axon-diameter - receptive field ratio. An inverse relationship exists between fiber diameter and receptive field size for diffusely organized receptors.

b. Controllable properties.

(1) Direct neural control. Central nervous system activity descending to the surface of the skin at the level of the peripheral receptor may control receptor properties by mechanical biasing. This may be accomplished by variation of the amount of blood filling specialized erectile tissue in the region of the receptor (106), or by action of small cutaneous muscles as, for example, the pilomotor system. Mechanical biasing would tend to influence:

- (a) Threshold, and thus threshold-diameter relationship,
- (b) frequency response, by damping and introduction of resonance peaks,
- (c) adaptation (that is the decrease in frequency of impulses propagated along the stem axon during the application of a constant intensity of stimulation to the peripheral receptor).

(2) Hormonal control. In addition to direct neural influence the central nervous system may release hormones either locally at the skin or generally into the circulation which will affect receptor properties (31). Humoral agents acting upon the nerve ending could conceivably affect threshold, frequency response, adaptation and accommodation. In addition

these agents by producing peripheral vaso-dilatation or vaso-constriction could also produce variations in mechanical biasing of the peripheral receptor.

(3) Antidromic control. In addition to neural activity initiated at central nervous system sites and distributed along orthodromic pathways, two possibilities exist for modification of the peripheral receptor by means of antidromic neural input. First: the dorsal root reflex (111) is an "echo" of neural input entering the central nervous system through the dorsal roots and leaving via these same and different dorsal roots. This "reflex" may effect the peripheral receptor by either colliding with impulses ascending from it or by depolarizing its terminals and thus introducing refractoriness. In addition an axon reflex may also occur. That is, a nerve impulse may arise from the peripheral receptor and be propagated from a collateral branch of the fiber leaving from that receptor without ever entering the central nervous system, and from this branch it may be propagated in turn into another portion of the peripheral receptor. This may also be effective in altering receptor properties by producing refractoriness. Primarily affected by this means would be frequency response and adaptation.

(4) Temperature. Changes in body temperature may affect the bias of temperature-sensitive units.

IV. Feedback Control of Cutaneous Information at First Synaptic Levels

The central nervous system structures operating upon information derived from the skin of the limbs and trunk, although similar in structure and general function, are different in detail from those mechanisms operating upon information derived from the skin of the face. The latter will not be considered in this report (118). Physiologic data are derived from studies on cat and monkey. Certain structures, particularly the lateral cervical nucleus have been found in cat and monkey, but not yet in man.

A. Lamina IV

All fibers entering the spinal cord in the dorsal roots (Fig. 11) have their cell bodies located in the dorsal root ganglia. As the dorsal root enters the spinal cord a reorganization of the distribution of its primary afferent fibers occurs. The large myelinated fibers carrying, as previously described, information from hair follicles and "touch spots", form a central core within the dorsal root which is surrounded on all sides by a thin layer of unmyelinated fibers (25).

When a single nerve enters the dorsal horn of the spinal cord, it projects directly to cells immediately adjacent to its zone of entry. These cells are excited, following a single synaptic delay, to produce a repetitive train of impulses (Fig. 7). Whatever inhibition may be produced by feedback mechanisms tends only to cut off further excitation and so shorten the evoked pulse train. At distances of 1-2 spinal

segments above and below the spinal segment of entry of the peripheral nerve, cells are no longer excited monosynaptically but only after traversing a pathway involving many synapses, and, weakly. Inhibition of cells excited by other nerves is now demonstrable. At levels three spinal segments above and below its level of entry, no excitation of cells, even polysynaptic, is seen following massive excitation of the peripheral nerve, and only inhibition of the activity produced by stimulation of other nerves may be observed. Thus, a single nerve may be considered to have a local sphere of influence, a strongly exciting core, a weakly exciting surround mixed with inhibition, and a final surround with inhibition only. Inhibition is probably conducted along polysynaptic pathways, as it may be present where no primary afferent fibers from the nerve stimulated can be found (109).

The central core of the dorsal root enters the spinal cord and those fibers derived from cutaneous receptors curve directly downward and form synaptic contacts with large cells in the gray matter of the dorsal horn of the spinal cord (Fig. 12). These large cells are arranged in a layer termed Lamina IV (Fig. 13) (99,100)* After giving off these synaptic

*The output of cells in Lamina IV of the spinal cord has been shown to be essential for the behavioral response reminiscent of "pain" perception in experimental animals (57). The velocity of conduction of impulses in the outflow tract of cells in Lamina IV has been demonstrated to be among the most rapidly conducting in the central nervous system (58 meters/sec) (110). (footnote con't on next page)

contacts the large cutaneous afferent fibers ascend dorsally in the spinal cord to Layers III and II of the dorsal horn where they end in tufts which lie among thickly packed dendrites and axons extending from a layer of small cells in this region (Fig. 14). The unmyelinated fibers of the dorsal root enter into the spinal cord and ascend rostrally about two to three spinal segments and caudally about two spinal segments in the medial division of a tract lying above Lamina I of the gray matter of the dorsal horn, termed the tract of Lissauer. Leaving this tract the unmyelinated fibers derived from the dorsal roots descend ventrally into Lamina II and III of the dorsal horn where it is thought that they make synaptic contacts with the small cells of this region and possibly with long dendrites of cells of the IVth Lamina, which ascend into this region (Fig. 15). The cells of the IVth Lamina project their axons in the lateral column of the spinal cord to cranial levels in a manner which shall be discussed below. The small cells of Lamina II and III project their axons into the lateral division of Lissauer's tract. This portion of Lissauer's tract, in contrast to the medial division, does not receive primary afferents. The axons of cells in laminae II and III

*(footnote from previous page continued). The information carried in this tract arrives after suitable modification at the somatosensory cortex a full 4 milliseconds ahead of information transmitted along other pathways (91). The rapid projection of the information carried from Lamina IV of the spinal cord into higher centers suggest that a function of these cells may be to provide the activation of regions of the central nervous system which have been demonstrated to be essential for certain aspects of consciousness (108).

then re-enter the grey matter of the spinal cord and terminate in Laminae II and III (115). Thus Laminae II and III provide feedback upon themselves via axonal projections in Lissauer's tract. The endings of these axonal projections from Lamina II and III may be upon one or all of the following three neural elements: (1) return feedback upon the cells of Laminae II and III; (2) termination upon long dendrites of the cells of Lamina IV which extend upward into Lamina II and III, as discussed; and (3) termination upon axons of the large myelinated and perhaps the smaller unmyelinated fibers entering the spinal cord from the dorsal roots.

The cells of Lamina IV, in addition to sending information up the spinal cord, may also send axonal projections ventrally to interstitial neurons whose axons project to motor cells in the ventral horn of the spinal cord and thus are involved in the generation of the polysynaptic reflex arc, which is involved in muscular activity. In addition to this projection, the cells of Lamina IV may feed back onto the cells of Laminae II and III, and also via interneurons onto their own dendrites, the dendrites of other Lamina IV cells, or onto the entering dorsal root fibers. Thus, cells of Lamina IV may control input to themselves.

It has been experimentally determined that when large myelinated fibers are stimulated alone, the feedback produced by activation of cells in Laminae II, III and IV, and acting upon the dorsal root fibers within the spinal cord, is

negative (26), and either tends to decrease the size of nerve action potentials progressing into the cord along the myelinated dorsal root axons, or to block nervous conduction along these axons entirely. On the other hand, it has been shown that when unmyelinated fibers are stimulated along with myelinated fibers in the dorsal roots, the overall effect is an increase in the output of cells in Lamina IV (80). The intimate mechanism for this facilitation is under experimental study, but may involve direct hyperpolarization of dorsal root axons by unmyelinated fibers or alternately an inhibition of those intermediary cells in Laminae II and III which tend to feed back negatively upon dorsal root axons. Thus, a complex balance of the effect of large and small fibers exists at the level of the very first synapse which interposes itself between the peripheral input and the processing of information ascending the spinal cord.

When a low intensity stimulus is delivered to the skin, both myelinated and unmyelinated fibers are active. As the stimulus is maintained on the skin, a balance is struck between the negative feedback from the cells of Lamina IV tending to decrease the dorsal root input, and the effect of the unmyelinated fibers tending to increase it. As the frequency of firing of unmyelinated fibers is limited, the potentiating effect of these fibers is limited, and the overall response is held within its bounds by the steadily increasing negative feedback effect from the myelinated group which continues to respond

over a wide range as the intensity of stimulation is raised.

In some disease states, the myelinated group appears to be damaged specifically. In these instances, the lightest peripheral stimulation results in a "flash" of pain. It may be the case in these conditions that negative feedback gain is reduced. In any event, the persistence of input-increasing tendencies provided by the unmyelinated fibers, coupled with the decrease in input-decreasing tendencies produced by an absence of myelinated fibers, results in an effective input increased above normal.

It is not to be supposed, however, that these local feedback interactions are alone in determining the overall output of the cells in Lamina IV of the dorsal horn. It has been determined that electrical stimulation of the brain stem over a wide region, in which a number of descending fiber systems was stimulated including fibers descending from the midbrain reticular formation, and from the vestibular system, an inhibition was demonstrated to be exerted directly upon the larger myelinated fibers entering the spinal cord in the dorsal roots (108). In addition, stimulation of the somatosensory cortex produces inhibition of dorsal root fibers entering the spinal cord, on the opposite side (6) (Fig. 17). Also, it has been shown that descending fibers from the somatosensory cortex make extensive synaptic contacts upon cells in Lamina IV (15,60,92). Excitation of these cells has never

been produced by stimulation of somatosensory cortex. It may be assumed that these synaptic contacts are inhibitory in nature. Thus the somatosensory cortex may influence the cells in Lamina IV of the spinal cord in two ways: first, by acting upon and inhibiting the input to them, and second, by inhibiting the cells themselves directly. The output of cells in Lamina IV of the spinal cord has been demonstrated to pass via a complex pathway to the contralateral somatosensory cortex, the same cortex which inhibits them. The output from Lamina IV follows a pathway (69,72) involving large cells in a special cluster just outside the gray matter of the spinal cord in the high neck region. These cells, forming the lateral cervical nucleus (12,41,42), in turn send their output across the spinal cord in the high neck region to the opposite thalamus (nucleus ventralis posterolateralis) (13,85). The cells from the thalamus send their output directly to the somatosensory cortex. The descending output of the somatosensory cortex is, in general (60,92), directed to those regions of the dorsal horn of the spinal cord from which they receive their afferent input. Thus a long negative feedback loop is established between the cells of the dorsal horn of the spinal cord and the cerebral cortex. In addition, the output of the cells of the dorsal horn of the spinal cord enters the reticular formation following its relay in the lateral cervical nucleus as discussed (85). From the reticular formation descending fibers return to the spinal cord, and, as has been described (108), inhibit the

entering afferent input. The great majority of Lamina IV cells studied respond both to hair movement and to stimulation of skin "touch spots" (116). They respond over a wide dynamic range, increasing their frequency with increase in the intensity of stimulation. It has been demonstrated that stimulation of a region on the skin which does not excite a specific cell in the Lamina IV of the gray matter of the spinal cord may act to inhibit it, probably through inhibition of the primary afferent input to it (108). This inhibitory skin region may be on the same limb, on the opposite limb, or in fact over a wide region of the body including the ears and nose. Thus, neurons in the dorsal horn of the spinal cord are controlled at local levels, as well as at intersegmental spinal levels, and, via complex feedback loops, at supra-spinal levels as well (Fig. 18).

B. Dorsal Column Nuclei (Fig. 19)

In parallel with the conduction system of cells in Lamina IV (8) is another conduction system, which in fact until recently had been considered to be the major system for the conduction of cutaneous information to the somatosensory cortex. As we have pointed out, the primary afferents entering the spinal cord from the dorsal roots form synapses with cells in Lamina IV. In addition, collaterals of these primary afferents ascend toward the head in the dorsal column which lies in the upper midportion of the spinal cord. The velocity

of conduction in the dorsal columns is about $2/3$ that in the root leading from Lamina IV. Upon reaching the high neck region, dorsal column fibers terminate in two large nuclei, the gracile nucleus, receiving input mainly from the hind limbs, and the cuneate nucleus (Fig. 20), receiving input mainly from the fore limbs. Until recent years it had been thought that the gracile and cuneate nuclei represent a simple relay of information ascending to the dorsal columns. This relay was thought to send its output across the medulla oblongata to the opposite side, then to the thalamus (nucleus ventralis posterolateralis) where it was then relayed again almost without modification to the somatosensory cortex. Investigation in recent years has demonstrated, however, a strong feedback both positive and negative, upon cells in the dorsal column nuclei from the somatosensory cortex and from the reticular formation (4,5,37,38,39,40,56,62,73,112,113).

The picture of function of the dorsal column nuclei is at present undergoing vigorous re-evaluation. An overview of its function may be given in terms of information presently available. As we have pointed out the cells in Lamina IV of the grey matter of the spinal cord, receive input both from myelinated and unmyelinated dorsal root afferents. In contrast, cells in the dorsal column nuclei receive input only from the largest myelinated afferent fibers (80). A considerable portion of these fibers have already made synaptic contact with cells in Lamina IV of the spinal cord, before

ascending in the dorsal columns to enter the dorsal column nuclei (110). Information reaching the dorsal column nuclei is in general of three kinds. First, a small number of neurons have been found to respond to manipulation of the joints (94). This is of significance in that in the classical neurological literature it had been assumed on clinical evidence that destruction of the dorsal columns produces a defect in localization of the position of the joints. However, this information was garnered from clinical cases in which much more of the spinal cord than the dorsal columns were damaged. Careful study of surgical patients in whom verification of section only of the dorsal columns was achieved at autopsy showed that during their lifetime no deficit in proprioception was obtainable and only a small deficit in two-point discrimination was noted (20). Thus, despite the complexity of the dorsal column nuclei to be discussed (43,70) the removal of all afferent input to these nuclei does not produce significant deficit in cutaneous sensation with testing techniques usually utilized (74). These techniques, however, are generally adjusted to reveal only threshold phenomena and cannot resolve sensory effects in time or in fact dynamic characteristics of the sensory process.

Second, the dorsal column nuclear cells were shown to receive information from hair movement (94) and, third, presumably from the "touch spots" previously described as peripheral afferent receptors, but now known to be part of

the "tylotrich sense organ" (74). This form of response to cutaneous stimulation differs from that shown by the cells in the spinocervical tract of the spinal cord in that all such cells respond both to movement of the hairs and to stimulation of the "touch spots", and the skin generally (116). Dorsal column nuclear cells, on the other hand, responded solely to one or the other modality (unimodal) (94,59). Several observers have confirmed the remarkable observation that only cells responding to movement of hairs possess receptive fields with inhibitory surrounds or other inhibitory regions on the skin (37,40,43,94). Cells which respond to the touch-temperature peripheral receptors do not possess receptive fields with afferent inhibitory surround. Concomitant with these observations it has been observed that fibers from the somatosensory cortex are inhibitory upon those cells whose receptive fields possess inhibitory surrounds (i.e. "hair responsive cells") and excitatory upon those cells which have no receptive fields with inhibitory surrounds (i.e. touch-temperature cells) (37,38,39). Cortical inhibition is found to be both pre- (4,5) and post- (39) synaptic, that is, it acts upon the input to the dorsal column nuclei and the cells of the nuclei themselves.

The cells responding to different forms of cutaneous peripheral receptor are not randomly distributed throughout the extent of the dorsal column nuclei, but, to the contrary, tend to be concentrated in specific regions (39). Hair-

sensitive cells lie mainly in the middle region of the nuclei and touch-temperature cells are found mainly in the rostral region of the nucleus, cells with smaller receptive fields in the middle and caudal region of the nucleus. Also cells in the rostral region of the nucleus tend to respond with longer delay than cells in the middle region of the nucleus. These delays were sufficiently long to suggest a multisynaptic organization within the dorsal column nuclei themselves (70,71).

Anatomical information (61) has demonstrated an interesting correlation between the modality of peripheral cutaneous receptor responded to and the anatomical organization of the cells in the dorsal column nuclei. Those cells responding to hair receptors were generally organized into tight clusters. Cells responding to touch receptors were generally organized in loose configurations. The fibers of the dorsal columns tended to contact the cells arranged in clusters in a very dense and precise synaptic organization, whereas cells arranged in a diffuse pattern were contacted by fibers ascending in the dorsal columns diffusely. It was demonstrated that cortical fibers are distributed predominantly to the loosely organized touch sensitive cells while much less so to the compactly organized cells arranged in clusters. Finally, it was shown that the cells arranged in clusters project directly into the major pathway leading to the nucleus ventralis lateralis of the thalamus whereas the cells loosely arranged, presumably receiving touch-temperature input,

projected elsewhere to reaches not clearly determined.

It thus appears that two forms of information processing exist within the dorsal column nuclei:

(1) A system wherein cells responding to stimulation of hairs, which have receptive fields with afferent inhibitory surrounds, are further inhibited by negative feedback from somatosensory cortical levels. These cells project directly to the nucleus ventralis posterolateralis of the thalamus.

(2) A system of cells receiving input from touch-temperature units, with fields which do not possess inhibitory surrounds, powerfully facilitated by somatosensory cortex. It projects to other regions than the nucleus ventralis posterolateralis of the thalamus.

In addition to long-loop cortical feedback to the dorsal column nuclei, both excitatory and inhibitory short-loop feedback to the cells of the nuclei was demonstrated (44). This short-loop feedback is of two types:

(1) Dorsal column nuclear cells send their axons into the major outflow tract of the dorsal column nuclei, the medial lemniscus, where these axons branch and return into the dorsal column nuclei. These recurrent collaterals may contact either the cells from which they arose or neighboring cells.

(2) In addition to direct contact of recurrent collaterals with their cells of origin or neighboring cells, a recurrent inhibition with a much slower time course than could be accounted for by direct contact of recurrent collaterals with cells bodies has been demonstrated. This is thought to involve a system of small cells upon which the recurrent collaterals terminate. These cells then contact neighboring larger cells which are responsible for the major portion of information transmission.

C. Summary of Interaction at First Synaptic Level

At this point it would be appropriate to summarize the basic similarities and differences between the two major cell systems for processing of information to be forwarded to the somatosensory cortex.

(a) Those cells of Lamina IV of the spinal cord forming the spino-cervical tract receive all sizes of primary afferent fibers, respond over a wide dynamic range, uniformly respond to both hair and touch temperature receptors, are under control by short positive and negative presynaptic feedback loops involving Laminae II and III, and are under control of long negative feedback loops acting, in the case of those derived from the reticular formation, presynaptically, and in the case of those arising from cerebral cortex both pre- and postsynaptically, always negative. All cells have receptive

fields with afferent inhibitory surrounds. Output from the cells of origin of the spino-cervical tract of the dorsal horn of the spinal cord ascends half again as fast (58.0 mps) to cervical levels as does information carried in the dorsal columns (38.3 mps).*

(b) Cells in the dorsal column nuclei receive input only from the largest unmyelinated fibers, respond over a narrow dynamic range, and respond either to hair or touch pressure receptors but not to both simultaneously. Cells responding to hair stimulation have receptive fields with afferent inhibitory surrounds. Cells responding to touch-temperature stimulation have receptive fields with no afferent inhibitory surrounds. Cells responding to hair stimulation receive negative feedback both pre- and postsynaptic from somatosensory cortex. Cells responding to hair stimulation project to the nucleus ventralis posterolateralis of the thalamus. Cells responding to touch temperature receptors do not project directly to the nucleus ventralis posterolateralis of the thalamus. All types of cells in the dorsal column nuclei receive recurrent inhibition, the precise nature of which is unknown, although it is thought, that along with the presence of recurrent synapses, producing inhibition that a system of small cells is also intermediate in producing afferent inhibition.

The previous discussion has dealt with the feedback loops and systems controlling the processing of information at

*mps = meters/sec

the first synaptic level in the spinal cord and in the dorsal column nuclei. Further sections will discuss the interaction of cutaneous information at higher neuronal levels, particularly the reticular formation, thalamus, and somatosensory cortex.

V. Control of Cutaneous Information at Thalamic Levels.

The properties of single units have been investigated in two portions of the thalamus of mammals --the ventral basal complex and the posterior nuclear group (Fig. 21). The ventral basal complex receives input primarily from the dorsal column nuclear pathway via the projection from the dorsal column nuclei, the medial lemniscus. The fibers from the medial lemniscus terminate uniformly in all sections of the ventral basal complex. In addition to this projection the ventral basal complex receives input from a system of fibers derived directly from the spinal cord. This system, termed the spino-thalamic tract, is prominent in primates and is less so as one descends the phylogenetic scale (77). In the cat, for example, no input arrives at the ventral basal complex directly from the lower spinal levels (13) and must pass through a synaptic relay at the lateral cervical nucleus (83,84,85,86,93).

A. Ventrobasal Complex

The input from spinal levels to the ventral basal complex of the thalamus distributes itself not uniformly in the ventral basal complex, as does the projection from the medial lemniscus, but rather it is arranged into islands or "clusters" of projections suggesting the term "archipelago" to the anatomical investigators who have studied this subject (77). Neurons of the ventral basal complex project to the cerebral

cortex, to Somatic Sensory Area I and perhaps to Somatic Sensory Area II, which will be discussed presently. Spinothalamic pathways in the primate project to a small extent directly to the posterior nuclear group (77)^(Fig. 22). The posterior nuclear group, in turn, has been determined to project to Somatic Sensory Area II.

The properties of neurons found in the ventral basal complex and in the posterior nuclear group differ markedly (88,96). The ventral basal complex is organized somatotopically, that is, discrete areas of the skin project to discrete regions of the ventral basal complex. Cells in the ventral basal complex respond uniquely to one form, and only one form, of peripheral cutaneous stimulation. Forty-two percent of neurons in the ventral basal complex are related to skin receptors, thirty-two percent to connective tissue or coverings of the bone, and twenty-six percent to the joints. The receptive fields of those units in the ventral basal complex responding to stimulation of the skin vary in size from 0.2 cm² on the fingers and toes to 20 cm² over the proximal parts of the limbs and body. This is similar in size to the receptive fields at first synaptic stations. The receptive field is organized with a central sensitive region and a peripherally less sensitive region (Fig. 23). Electrical stimulation at the most sensitive region produces beat for beat following of the stimulus by the ventral basal complex neuron, even to frequencies as high as 400/second. At the periphery

of the receptive field, following is less precise and fails at frequencies above 100/second. Two characteristic properties of neurons in the ventral basal complex are: (1) their rapid onset transient and dead-beat ending and (2) their inhibition from a relatively discrete region surrounding the receptive field, provided the stimulus for producing such inhibition leads the stimulus to the receptive field by about 3 milliseconds. Such afferent inhibition (Fig. 24), which has been described for the spinocervical tract and for hair sensitive cells in the dorsal column nuclei (94), is eliminated by anesthesia (106) as is afferent inhibition in the latter areas (88). Studies of thalamic units have not yet differentiated response to cutaneous stimulation into hair and touch modalities. Neither have there been studies relating to the effect of stimulation of cerebral cortex and reticular formation upon responsiveness of single thalamic neurons to peripheral cutaneous stimulation. However, it appears probable that neurons in the ventral basal complex reproduce in their properties, those of the spinocervical tract as well as those of the hair sensitive cells in the dorsal column nuclei. No units without afferent inhibition were found, and so it must be supposed that no line from "touch" units exists to the ventral basal thalamus.

B. Posterior Nuclear Group

Neurons of the posterior nuclear group of the thalamus respond to stimulation of very large peripheral receptive fields some of which cover the body surface (96) (Fig. 25). No somatotopic representation exists in the posterior nuclear group. In contrast to the cells of the ventral basal nuclear complex, cells of the posterior group may respond to several modalities of stimulation, for example, some respond to light touch, increase their response when tissue damaging stimuli are delivered, and may also respond to auditory stimuli. In contrast to the discrete form of afferent inhibition found for the ventral basal nuclear cells, cells in the posterior group display a complex topography of excitation and inhibition. Cells activated from one side of the body may be inhibited from the other side. Other cells activated from more restricted fields may be inhibited from almost all the rest of the body. A characteristic response of the cells of the posterior nuclear group is a slow rise in frequency followed by a prolonged after-discharge following cessation of the stimulus. This pattern is distinct from the rapid onset transient and dead-beat endings of the discharges of cells of the ventral basal nuclear complex.

Comparison of the properties of the ventral basal nuclear group with that of the posterior nuclear group suggest that powerful feedback inhibition is evoked at some point in the

pathway leading to the ventral basal nucleus, whereas a weaker form of feedback inhibition or, in fact, feedback facilitation is evoked upon the cells of the posterior nuclear complex.

It seems from these response properties that the posterior nuclear group may be associated with those processes relating to the mediation of pain, whereas neurons in the ventral basal complex seem to be unrelated to these processes. Nevertheless the projection of the spinocervical tract upon the ventral basal complex and the necessity for the integrity of either the spinocervical tract in the cat or the anterolateral spinothalamic tract in man for the pain experience following vigorous cutaneous stimulation suggests that mechanisms for the mediation of such pain experience are located in the ventral basal nuclear complex as well.

VI. Control of Cutaneous Information at the Cortical Level

A. Cortical Subdivisions

On the surface of the cerebral cortex (Fig. 26) it is possible to record evoked potentials of the order of several millivolts when the skin is stimulated. These potentials are sharply localized to cortical regions deriving input from specific skin areas. Two complete maps of the cutaneous surface are present on the surface of the cerebral cortex (Fig. 27). The first and more precise map is located in the region of the cerebral cortex called Somatic Sensory Area I. (Fig. 28). This region receives input from the opposite half of the body only. The receptive field properties of cortical cells in Somatic Sensory Area I are similar to properties of cells found in the ventral basal complex (88). No marked difference in average size of receptive field is found when the cortex and thalamus are compared. Somatic Sensory Area II, located somewhat more laterally on the surface of the cerebral cortex than Somatic Sensory Area I, receives input from both sides of the body (88).

B. Somatic Sensory Area I (88)

The vast majority of neurons in Somatic Sensory Area I are characterized by unimodal response, some cells responding

to joint movement but most possessing either low threshold hair-sensitive or touch-sensitive sharply delimited receptive fields recorded in the dorsal horn of the spinal cord, the dorsal column nuclei and the thalamus; i.e., distal fields smaller in area, proximal fields larger. Under light anesthesia, afferent inhibition is evident. As in the ventrobasal nucleus of the thalamus, the receptive fields of the cortical neurons show a central sensitive region and a more peripheral, less sensitive, region. Units stimulated from the central more sensitive region follow frequencies of stimulation as high as 400/sec whereas more peripheral stimulation results in a decrease in the ability to follow high frequency input. As is usual in somatic sensory systems, a single peripheral shock results in a repetitive train of spikes recorded from a single unit in the cerebral cortex with as many as ten action potentials occurring in rapid succession at rates up to 1000/sec, following a single peripheral stimulus. Neurons responding to the same modality are located in columns of cells responding similarly. It is found that the cells responding from the center of the receptive field have a shorter response latency.

C. Somatic Sensory Area II (14)

In Somatic Sensory Area II eight percent of cells are found to be similar in properties to cells recorded in Somatic Sensory Area I, with the exception that in Somatic Sensory Area I, many cells are found activated by gentle rotation of

joints. However, the Somatic Sensory Area II contains another population of cells which, in contrast to the majority of cells in Somatic Sensory Area I, have neither restricted receptive fields nor do they respond unimodally. These cells are characterized by the fact that their receptive fields are frequently large and bilateral, that they may be activated by light mechanical stimulation in one part of the field and only by noxious stimulation in others, that some are activated only by noxious stimuli and others may even be influenced by sound stimuli. Interaction between somatic sensory and visceral stimuli upon these cells have been observed and measured (3). A small number of cells with the properties described have been reported to be present in Somatic Sensory Area I.

As may be noticed the non-place-specific non-modality-specific cells in Somatic Sensory Area II resemble closely in properties those cells recorded from in the posterior nuclear thalamic group and since evidence is present showing projection of the posterior nuclear thalamic group onto Somatic Sensory Area II, it may be concluded that Somatic Sensory Area II receives input from the posterior nuclear group essentially unchanged.

For many years it has been postulated that information similar to that recorded in Somatic Sensory Area II reaches cerebral levels from the periphery via the "spinothalamic tract". Thus some authors have postulated that the cells in

the posterior nuclear thalamic group receive input by way of the "spinothalamic tract" and not from the "medial lemniscus" (96). However, it has been shown that when the only tracts available for transmission into the central nervous system are the dorsal columns, which are known to project, via the dorsal column nuclei, into the medial lemniscus, units in Somatic Sensory Area II are recorded which show nonmodality-specific and non-place-specific properties (7). If, indeed, these cortical units receive their input from the posterior nuclear group, it must be postulated that the posterior nuclear group and thus Somatic Sensory Area II receives input via the dorsal columns as well as the postulated "spinothalamic tract".

The results just discussed have been those relating to the response of cells in the cerebral cortex to peripheral cutaneous stimulation. What processes intervene between the response of these cells and the subsequent production of a descending output by the cerebral cortex is unknown. It is also not clear whether experiments involving electrical stimulation of the cortex producing either facilitation or inhibition of neurons in the spinal cord or in the dorsal column nuclei are, in fact, related to the characteristic responses of single cells reported as input to the cerebral cortex, and presumably responsible for this output. In any event, as has been previously discussed, electrical stimulation of the cerebral cortex does produce at least

upon units in the dorsal column nuclei either facilitation or inhibition depending upon the type of unit recorded from. With respect to cells in Lamina IV of the spinal cord, the putative cells of origin of the spinothalamic tract, only a slight, long lasting, inhibition has been demonstrated following stimulation of the sensorimotor cortex* (68).

It must not be supposed that the cortex sends projections directly to the spinal cord only. While not considered in detail in this report, it is known that the cerebral cortex

* Note: the term "sensori-motor" cortex is ambiguous. A portion of the cerebral cortex was found from the earliest investigations in neurophysiology to produce movements of the extremities when stimulated with appropriate currents. Few other cerebral cortical regions manifested this property. Accordingly this cortical region was termed the "motor" cortex. Subsequently it was shown anatomically, that input was received by the "motor cortex" from the thalamus, and electrophysiologically it was demonstrated that input was also received from peripheral cutaneous stimulation. Thus the term "sensori-motor" cortex was coined. In reading the literature, it is unclear when the term "sensori-motor" cortex is used, whether in fact the cortex previously called "motor" was stimulated or the cortex called "sensory" was also stimulated, that is, Somatic Sensory Area I and Somatic Sensory Area II. The distinction is important to make since it has recently been demonstrated that descending projections from the old "motor" cortex descend to the spinal cord and terminate at cell groups ventral to the spinal cord to those groups which are in Lamina IV. Contrarywise projections from the true "sensory" cortex, Somatic Sensory Area I, have been found to form dense synaptic connections with cells in Lamina IV. The failure to find excitation or inhibition in cells in the general region of Lamina IV by stimulation of the "sensori-motor" cortex may thus not be indicative of the presence or absence of feedback from the true sensory areas of the cerebral cortex to the cells of Lamina IV.

sends projections to, among other regions, the thalamus and the region in the brainstem termed the reticular formation. These regions will not be discussed extensively except to point out that stimulation of the reticular formation electrically in the unanesthetized animal produces a change in the spontaneous electroencephalogram of the animal, that is, a decrease in voltage and an increase in frequency, similar to that change produced when the animal is aroused by a "natural stimulation" or "spontaneously". Stimulation of the reticular formation facilitates various motor reflexes at the same time as it produces electroencephalographic "arousal". Continuing stimulation of the reticular formation with the cortex intact results, eventually, in a decrease of the capability of the reticular formation to facilitate motor reflexes, suggesting that continued arousal of the cerebral cortex results in inhibition of the reticular formation (50). That is, cortical activation produces negative feedback upon reticular activity. This is confirmed when the cerebral cortex is removed and the reticular formation is stimulated continuously in which case the facilitation of motor responses is not decreased. It is at present unclear what effect, if any, the electroencephalographic manifestations of arousal have upon single unit responses in the cortex recorded following the stimulation of periphery. When such responses to "natural stimuli" are studied with gross electrodes, they appear to be inhibited during arousal. However,

when these responses are evoked by massive electrical stimulation of the peripheral nerve, they tend to be facilitated by such electroencephalographic arousal (11). The precise interaction between the electroencephalographic arousal process and the response of individual units in the cerebral cortex is not known at the present time. However, it can be said that a complex interrelationship exists between the reticular formation and the cerebral cortex wherein stimulation of the so-called facilitatory reticular formation results in electro-cortical arousal followed, after a delay, by a production of negative feedback which in turn inhibits the reticular formation.

In addition, to what has been discussed it should be pointed out that regions in the reticular formation, particularly near those portions of the brainstem which receive input from the viscera, when stimulated electrically produce slowing of the electroencephalogram suggestive of those changes which occur in drowsiness and sleep (21). These regions in the reticular formation appear to be those excited when the viscera are distended with food. Undoubtedly this form of interaction with the cerebral cortex should influence the processing of sensory information at cortical levels. However, no definitive study is as yet available on this subject.

Complex interaction between the input from the viscera and the cutaneous input has been found, not only in the classical Somatic Sensory Cortices I and II, but in regions

of the cerebral cortex termed Association Cortex (3). To add to the complexity of the problem, it has been shown that stimulation of the head of the caudate nucleus, a region in the brain which recently has been found to receive a complex cutaneous input, inhibits potentials evoked in the association cortex by peripheral stimulation (2, 58).

Other regions in the brain which have been associated with cutaneous sensation include a nucleus of the thalamus, the centrum medianum. This nucleus appears to receive its major input from a region in the lower brainstem reticular formation, the nucleus gigantocellularis (2). Removal of the centrum medianum in man has been found to be associated with a decrease in the reaction to painful stimulation with a retention of a clear state of consciousness. It is of interest to note that a nucleus in the reticular formation, the nucleus reticularis gigantocellularis, is located precisely in that region where previous investigation has suggested a projection from the lateral cervical nucleus. As we have seen previously the lateral cervical nucleus is the cervical relay for the dorso-lateral spinal cutaneous afferent pathway which takes its origin in the cells of Lamina IV of the spinal cord. This pathway at least in its spinal part, has been associated with the perception of pain in animals (57). It may be that the effect of the spinocervical pathway upon pain is ultimately upon the centrum medianum of the thalamus via nucleus reticularis gigantocellularis. Support for this concept has been gained

recently by studies in which the pathway in the spinal cord for the input to the nucleus reticularis gigantocellularis were investigated. It was found that stimulation of the dorsal column produced impulses which ascended in the lateral portion of the spinal cord, presumably after synapse in the dorsal horn. As we have previously discussed, the dorsal columns send collaterals from a portion of their fiber content into the dorsal horn synapsing upon cells of origin of the spinocervical tract; thus the spinocervical tract may indeed be the major tract transmitting information to the nucleus-gigantocellularis-centrum medianum system.

Finally it must be made clear that a large number of structures not discussed in this report receive input from the skin after many synaptic contacts. This, of course, is to be expected in that a very large portion of the apparatus of the brain is devoted to processing of cutaneous information. At what stage in the processing of such information the neural input is being studied when it is looked at experimentally in a particular neural structure, or indeed, if it is "input" can of course not be decided until a comprehensive theory of the processing of central nervous system information is evolved (Fig. 29).

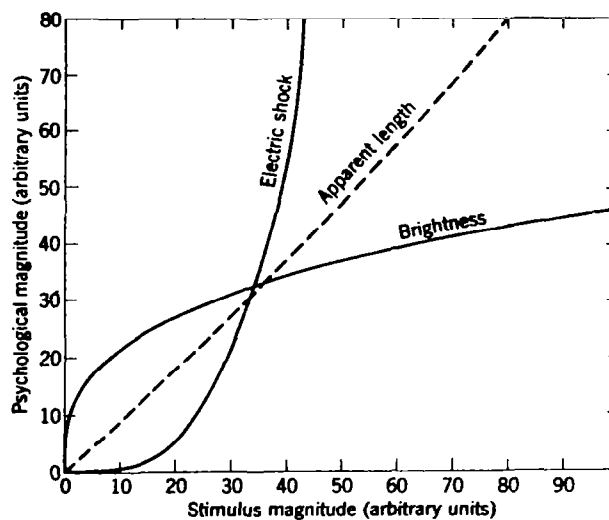
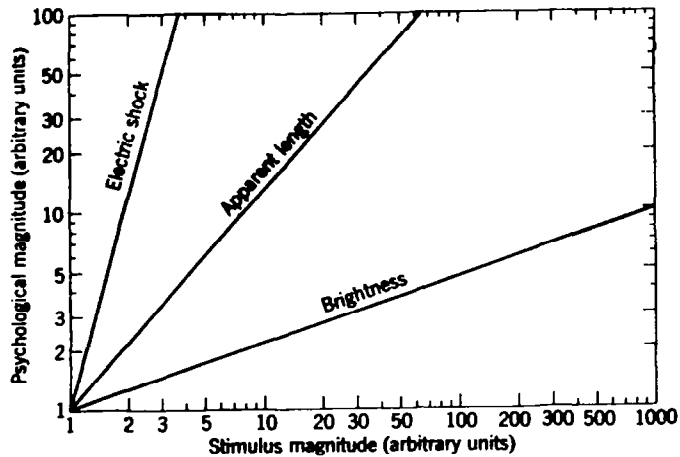


Figure 1. (top figure) Scales of apparent magnitude for three prothetic continua plotted in log-log coordinates. The slope of the line corresponds to the exponent of the power function governing the growth of the psychological magnitude.

(lower figure) In linear coordinates the subjective magnitude functions slope upward or downward depending on whether the power-function exponent is greater or less than 1.0. (105)

Continuum	Exponent	Stimulus conditions
Loudness	0.6	Binaural
Loudness	0.54	Monaural
Brightness	0.33	5° target—dark-adapted eye
Brightness	0.5	Point source—dark-adapted eye
Lightness	1.2	Reflectance of gray papers
Smell	0.55	Coffee odor
Smell	0.6	Heptane
Taste	0.8	Saccharine
Taste	1.3	Sucrose
Taste	1.3	Salt
Temperature	1.0	Cold—on arm
Temperature	1.6	Warmth—on arm
Vibration	0.95	60 cps—on finger
Vibration	0.6	250 cps—on finger
Duration	1.1	White-noise stimulus
Repetition rate	1.0	Light, sound, touch, and shocks
Finger span	1.3	Thickness of wood blocks
Pressure on palm	1.1	Static force on skin
Heaviness	1.45	Lifted weights
Force of handgrip	1.7	Precision hand dynamometer
Autophonic level	1.1	Sound pressure of vocalization
Electric shock	3.5	60 cps, through fingers

Figure 2. Representative exponents of the power functions relating psychological magnitude to stimulus magnitude on prothetic continua. (105)

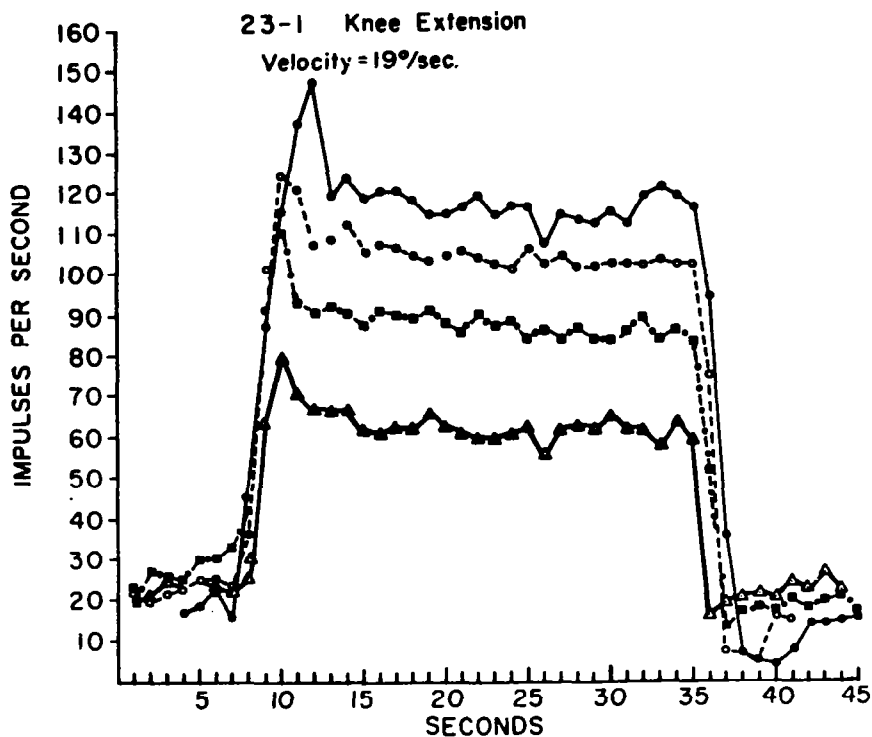


Figure 3. The results obtained during the "one speed to several angles" experiment for neuron 23-1, a ventrobasal thalamic cell driven by extension of the contralateral knee. For each angle, five trials were made, and the data listed in terms of impulses per 200-msec. counting period. The five lists were then oriented correctly in time, averaged and summed for each second, and finally plotted, as shown here, as impulses per second. The knee was rotated for a position well outside the excitatory angle to true joint angles of, from above downward, 180° , 125° , 100° , and 80° respectively. Movements towards extension were begun at the 7th sec., and those towards flexion at the 35th sec. (89)

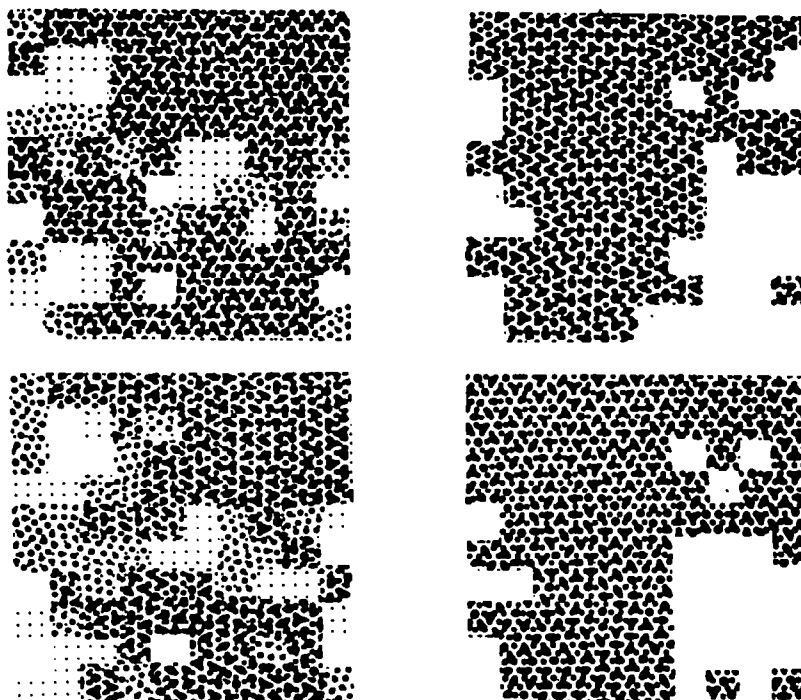


Figure 4. Fluctuation of cold sensitivity: changes at the boundaries of sensory fields. Left, two successive maps. Right, two maps from another subject based on reports of "cold" (stipple) or "no cold" (white areas). Each map was completed in approximately 20 minutes; 20-minute interval between two successive maps. (78)

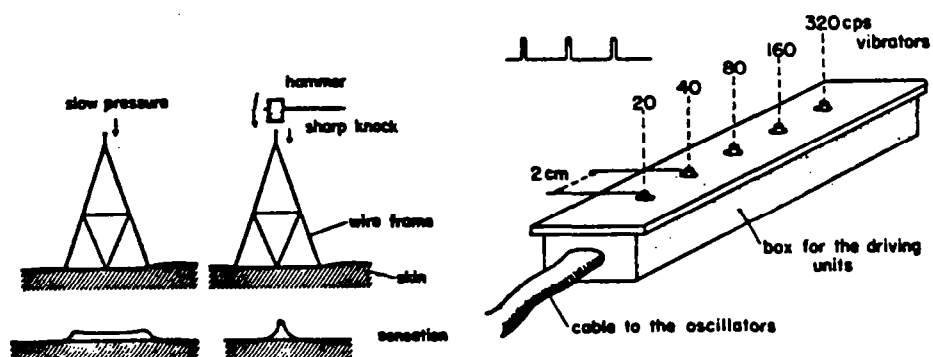


Figure 5. (Left hand figure) Pressure applied slowly to the surface of the skin produces a sensation as large as the size of the object. But if an object is tapped against the skin, a sensation is observed only in the center of the object.

(Right hand figure) Series of vibrators, 2 cm apart and increasing in frequency from left to right. The vibrations consisted of a series of pulses similar to one shown in the left-hand corner. The unit was placed along the lower arm. (10)

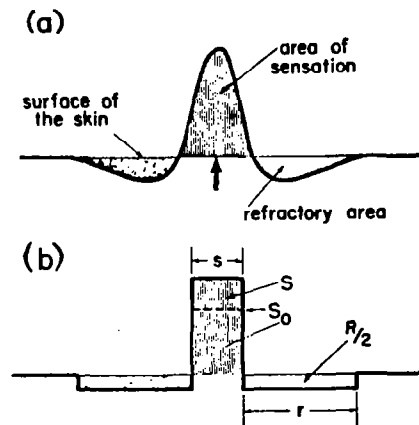


Figure 6. The pressure of a single point on the surface of the skin produces (a) an area of sensation and around it a refractory area in which a neighboring stimulus is inhibited; in (b) the pattern is simplified to a rectangular shape. S is the surface of the sensation area and $R/2$ the surface of the refractory area on one side. S_0 indicates the height of the sensation area for which $S_0 = R$. (10)

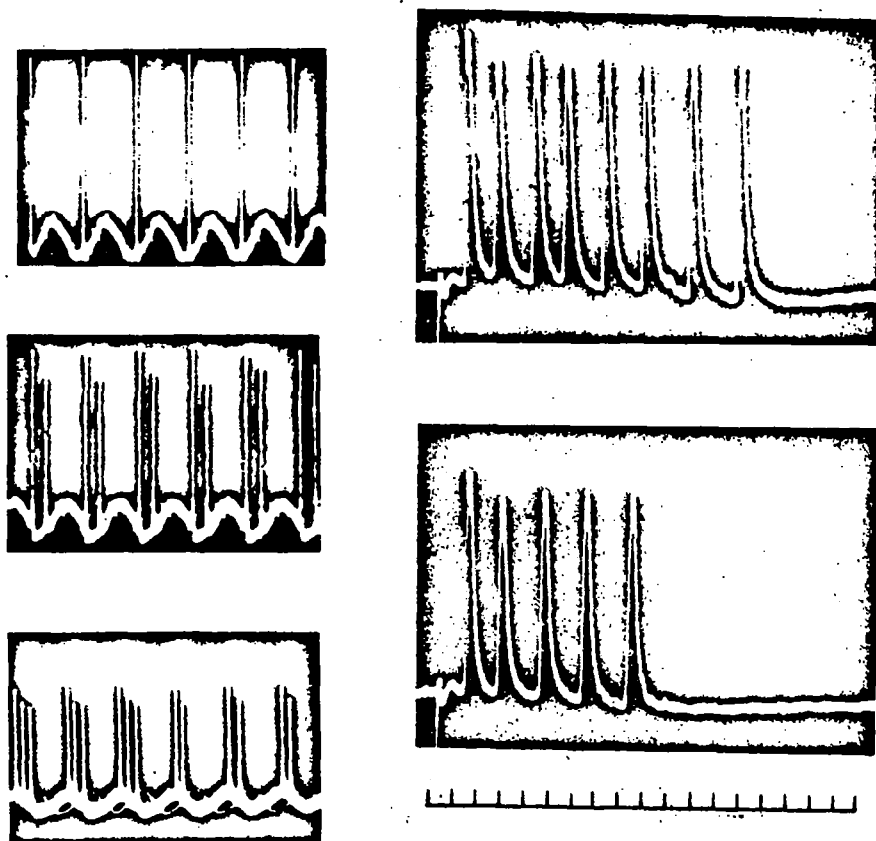


Figure 7. (Left hand figure) Impulses recorded during light vibration of cat skin; recording by intracellular microelectrode. Metal plate vibrating at 60 cps; variations of the base line indicate variations of pressure on the skin. The top picture was recorded from a large afferent axon in the dorsal root and shows a single response to each increase of pressure on the skin. The lower two pictures were recorded from within primary central cells and show repetitive responses to each pressure increase. In the lower record, it will be noted that the repetitive response varies between 2 and 3 impulses per burst. Time is given by the base-line variations at 60 cps.

(Right hand figure) The effect of skin vibration on evoked response in dorsal horn neuron. Light vibration at 60 cps applied to the receptive skin field of the neuron; neuron responded to each increase of pressure. Stimulus applied to intact dorsal root. Recording from within a central cell by glass microelectrode. The upper record shows the repetitive discharge which followed a decrease in length of the repetitive burst when the same shock is applied during the application of vibration to the skin. The vibration itself was producing intermittent excitation of the cell. Time is in microseconds. (117)

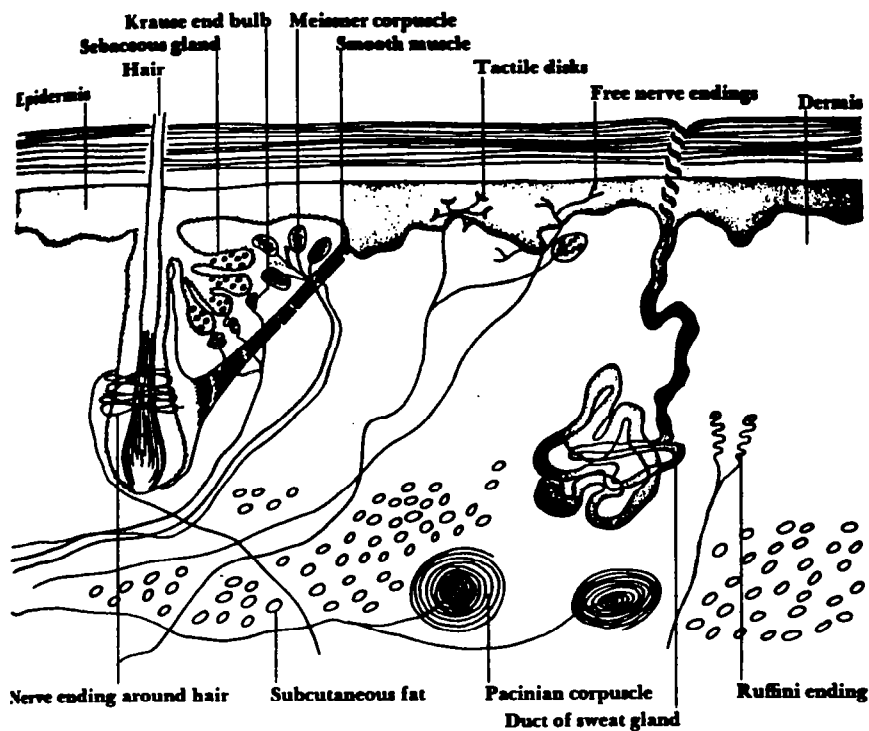


Figure 8. Schematic drawing of a cross section of the human skin. (Adapted from H.H. Woolard, G. Weddell, and J.A. Harpman. Observations on the neurohistological basis of cutaneous pain. *J. Anat.*, 74:413-440, 1940 and E. Gardner. Fundamentals of Neurology. Philadelphia: Saunders, 1947, p. 111.) (81)

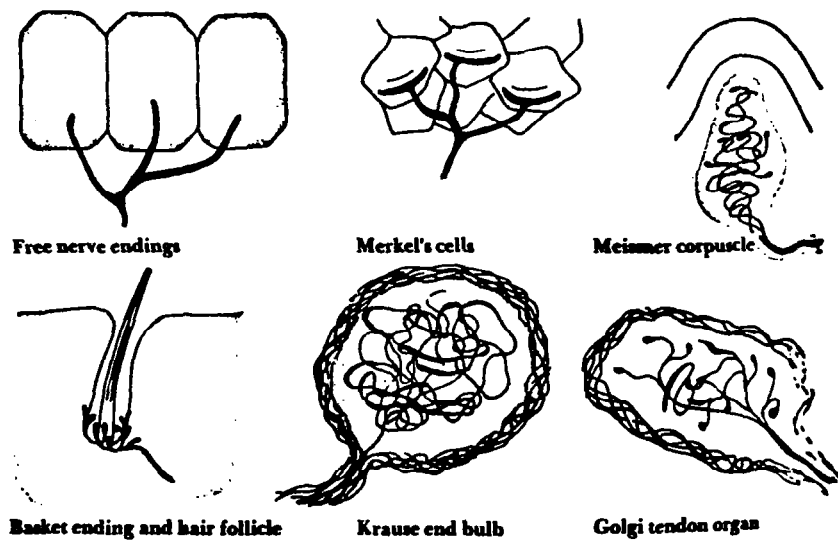


Figure 9. Diagrams of the principal receptors of the skin. (Adapted from J.F. Fulton. Physiology of the Nervous System. 2nd ed., Fair Lawn, N.H.: Oxford University Press, 1943, p.3 by permission of the publishers.) (81)

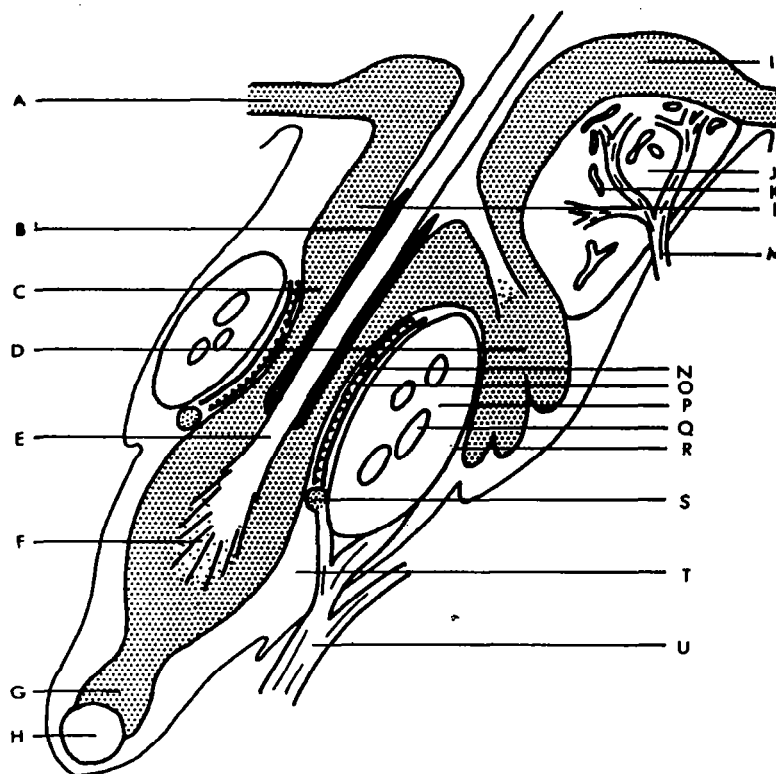


Figure 10. A generalized hair follicle. Generalized tylotrich follicle in the quiescent phase of hair growth. The names of structures generally characteristic of the tylotrich follicle are keyed with letters at the right of the drawing. Structures characteristic of both tylotrich follicles and of other pelage hair follicles are keyed at the left. The corresponding names are listed below. (106)

A, epidermis
 B, internal root sheath
 C, external root sheath
 D, sebaceous gland
 E, hair
 F, hair club
 G, germ of follicle
 H, dermal papilla
 I, Haarscheibe
 J, specialized region of dermis below Haarscheibe
 K, capillary within specialized region of dermis
 L, thickened area of external root sheath

M, branch of large nerve, innervating Haarscheibe and specialized region of dermis
 N, bilaminar arrangement of nerve fibers
 O, band of smooth muscle-like cells
 P, annulus
 Q, capillary within annulus
 R, annulus sheath
 S, annular nerve
 T, connective tissue capsule
 U, large nerve innervating annular complex, Haarscheibe, and specialized region of dermis.

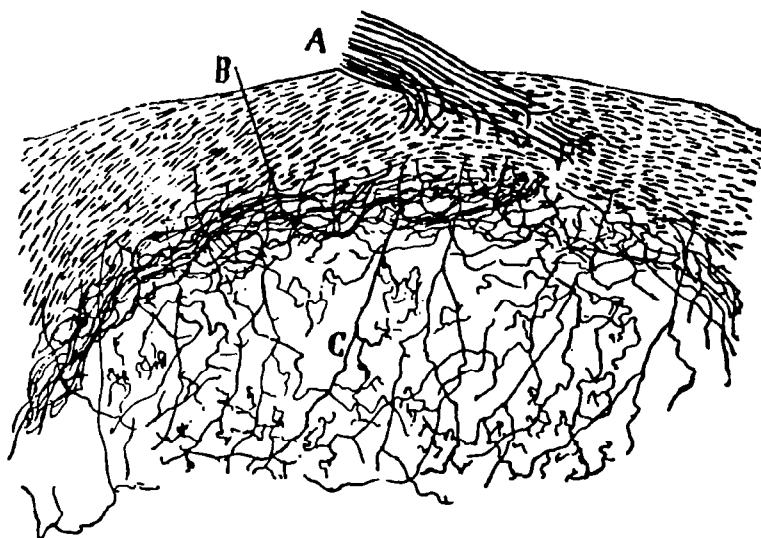


Figure 11. Drawing from Ramón y Cajal (1952) to show the collaterals of small dorsal root fibers which form a plexus in lamina I (B) and penetrate radially into laminae II and III (C). The original legend (Ramón y Cajal, 1952, Fig. 120) is: Coupe transversale d'une partie du cordon postérieur et de la substance de Rolando, dans la moelle lombaire; chat nouveau-né. Méthode de Golgi. (A) racine postérieure; (B) plexus marginal de collatérales; (C) collatérales fines, allant à la substance de Rolando. (104)

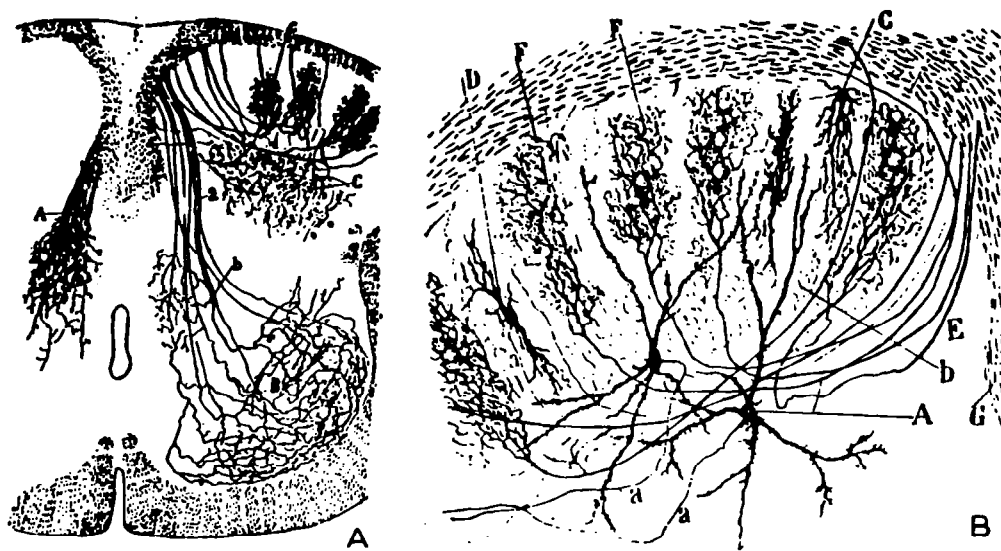


Figure 12. Two drawings from Ramón y Cajal (1952) to show four terminal areas in the gray matter of the spinal cord (A). These are (c) arborizations of large dorsal root fibers in laminae II and III, (C) plexus and terminals in lamina IV, (b) terminals in the intermediate gray nucleus, (B) terminals in the somatic motor nuclei. An enlarged drawing (B) shows the synaptic compartments of the arborizations in laminae II and III, small neurons of these laminae, and large neurons of lamina IV.

The original legend (Ramón y Cajal, 1952, Fig. 113) is: Principales collatérales sensitives, chez le rat nouveau-né. Méthode de Golgi. (A) collatérales du noyau gris intermédiaire; (B) arborisations embrassant les noyaux moteurs; (C) ramifications étendues dans la tête de la corne postérieure; (a) faisceau sensitivo-moteur (b) collatérale d'une des fibres destinées au noyau gris intermédiaire; (c) collatérales profondes de la substance de Rolando.

The original legend (Ramón y Cajal, 1952, Fig. 121) is: Coupe transversale de la substance de Rolando, dans la moelle cervicale; chat nouveau-né. Méthode de Golgi. (A) cellules de la tête de la corne postérieure; (B, C, D) cellules de la substance de Rolando; (E) collatérales grosses ou profondes de cette substance; (F) arborisations nerveuses terminales provenant des collatérales profondes; (a) cylindre-axe; (b) arborisations nerveuses longitudinales du sommet de la corne postérieure.

(104)

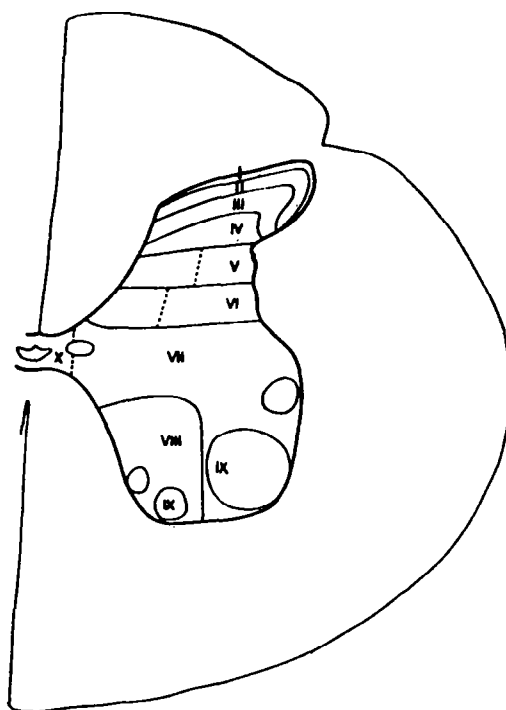


Figure 13. Schematic drawing of the lamination of the spinal cord grey matter of the 5th lumbar segment in the adult cat. (100)

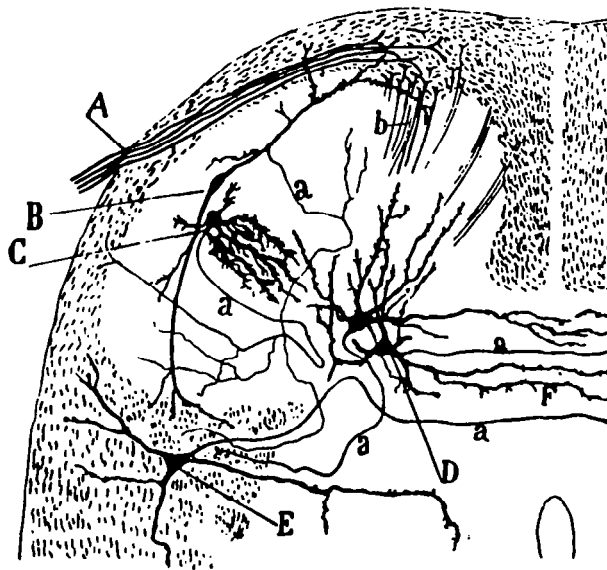


Figure 14. Cells of the head of the dorsal horn. Note two types of cells (B and C) in the substantia gelatinosa Rolandi. Golgi technique. Hen embryo, 15 days incubation (Cajal, 1909). (100)

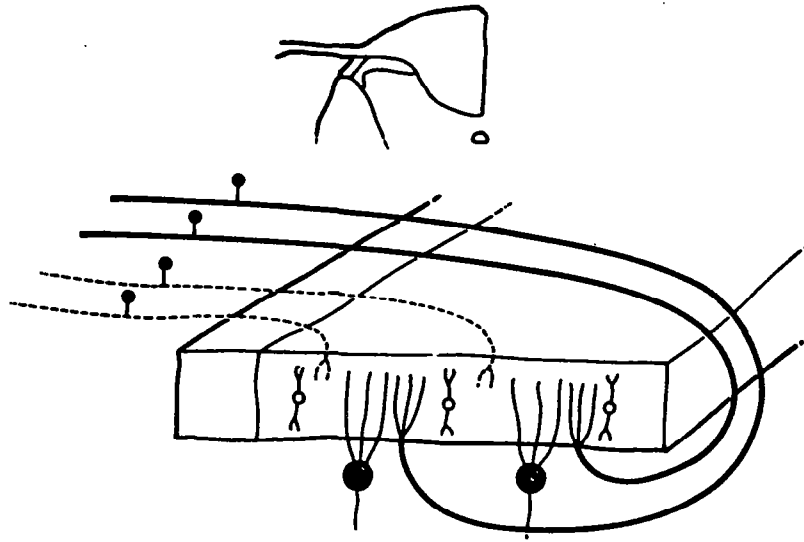


Figure 15. The upper drawing shows a cross section of the dorsal quadrant of cat lumbar cord. The lower diagram shows the three main components of the cutaneous afferent system in the upper dorsal horn. The large diameter cutaneous peripheral fibers are shown as thick lines running from dorsal root to terminate in the region of the substantia gelatinosa. The finer peripheral fibers are shown as dashed lines running directly into the same region. The large cells of lamina IV of Rexed on which cutaneous afferents terminate are shown as large black spheres with their dendrites extending dorsally. The small cells represent the cells of substantia gelatinosa. Their axons are not shown, but interconnect the cells of substantia gelatinosa and also run in the Lissauer tract which is shown as the most lateral structure. (115)

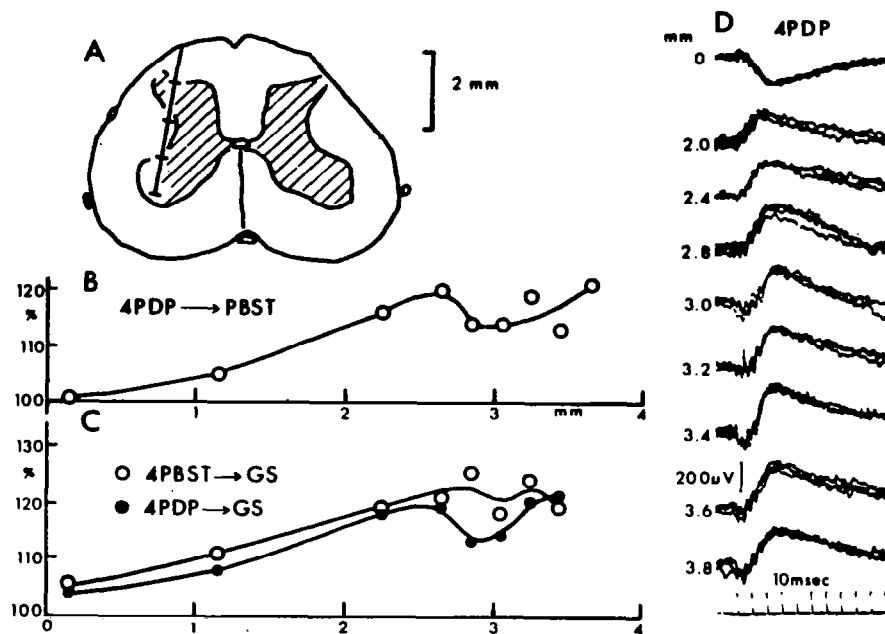


Figure 16. Locations of depolarizing foci on primary afferent fibers. B and C are excitability increases plotted against the depth in mm along a microelectrode track rather more medial than that in A, where the depths in mm are marked. The conditioning stimuli and the fibre terminals on which they were tested are indicated for each series. In D the series of focal potential records with three or four superimposed traces were at the indicated depths, in mm, along the track shown in A and were evoked by four PDP Group I volleys at 300/sec. A is a traced enlargement of the section through the spinal cord with the microelectrode in situ. Upward deflections in D signal negativity. Same potential and time scales for all traces.

(26)

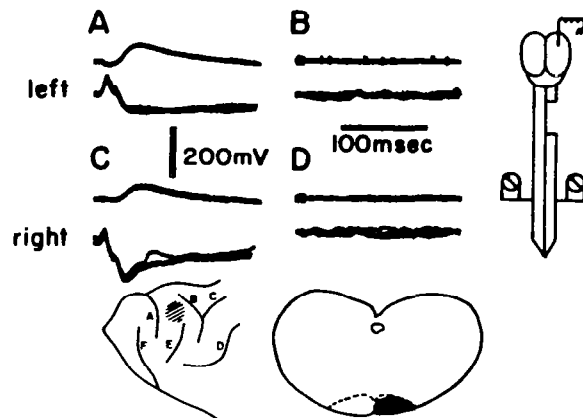


Figure 17. Primary afferent depolarization evoked from the sensorimotor cortex. The dorsal root potentials in A and C (upper traces) were evoked from the right sensorimotor cortex and recorded from the most caudal dorsal rootlet in L7 on the left and right side as indicated. The lower traces were all recorded with one electrode placed on the dorsal column and negativity is signalled by an upwards deflection. The corresponding records B and D were obtained after section of the right pyramid (cf. lower right diagram). The shaded area in the lower left diagram is the region from which actions could be obtained at threshold stimulation. (68)

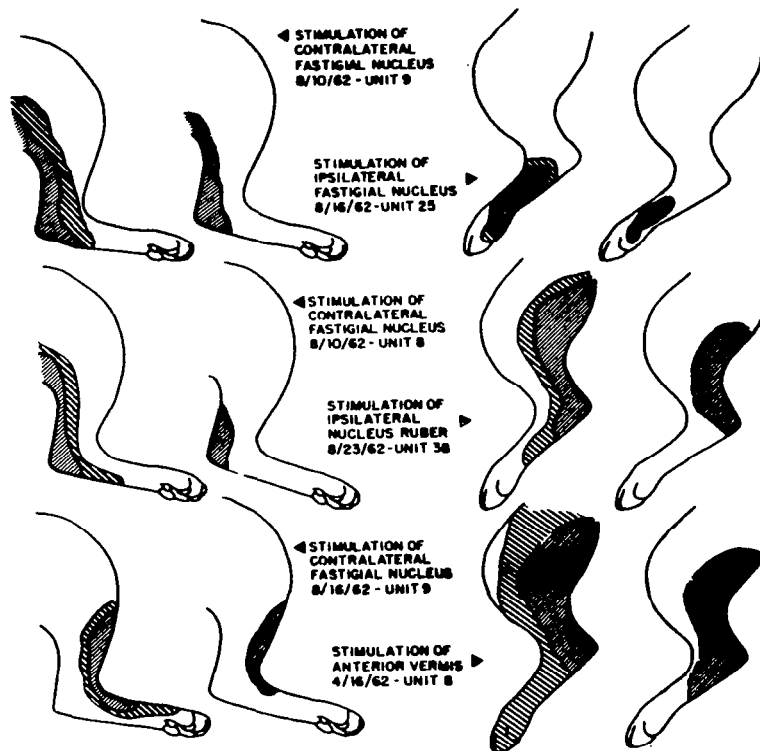


Figure 18. Tactile receptive field constriction of some units in the spinocervical tract during supraspinal stimulation. Periphery of excitatory receptive field, which was generally less sensitive, is now completely insensitive, whereas the "central" portion of the receptive field still responds, although weakly. A general constriction of the natural stimulus excitatory receptive field is evident. Fine crosshatching = excitatory receptive field coarse crosshatching = periphery of the field. (108)

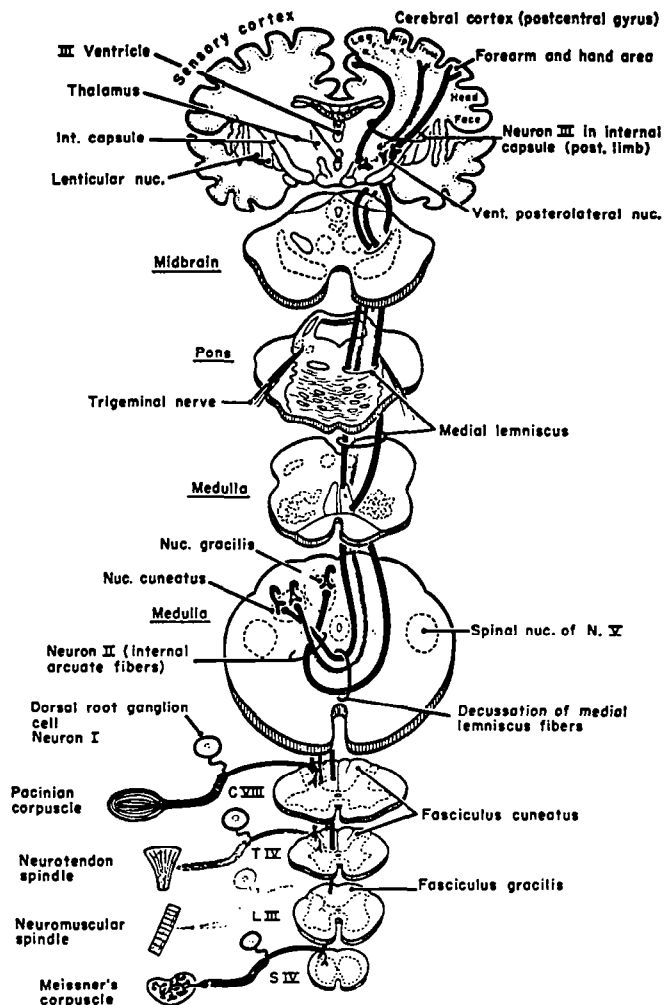


Figure 19. Classical diagram of dorsal column-dorsal-column-nuclei medial-lemniscus system, in which the dorsal column nuclei are depicted as monosynaptic and no feedback interactions are shown. The precise nature of the sensory receptors affording input to this system is not known experimentally despite their presence in the diagram. Nevertheless, the diagram serves to place the relationship of significant structures into perspective. The spino-cervical tract is not depicted. It has not yet been established for man, but has been shown to exist in primates. (From Truex, R.C., and Carpenter, M.B., Strong and Elwyn's Human Neuroanatomy, Fifth Edition, Williams and Wilkins, Baltimore, 1964, page 206).

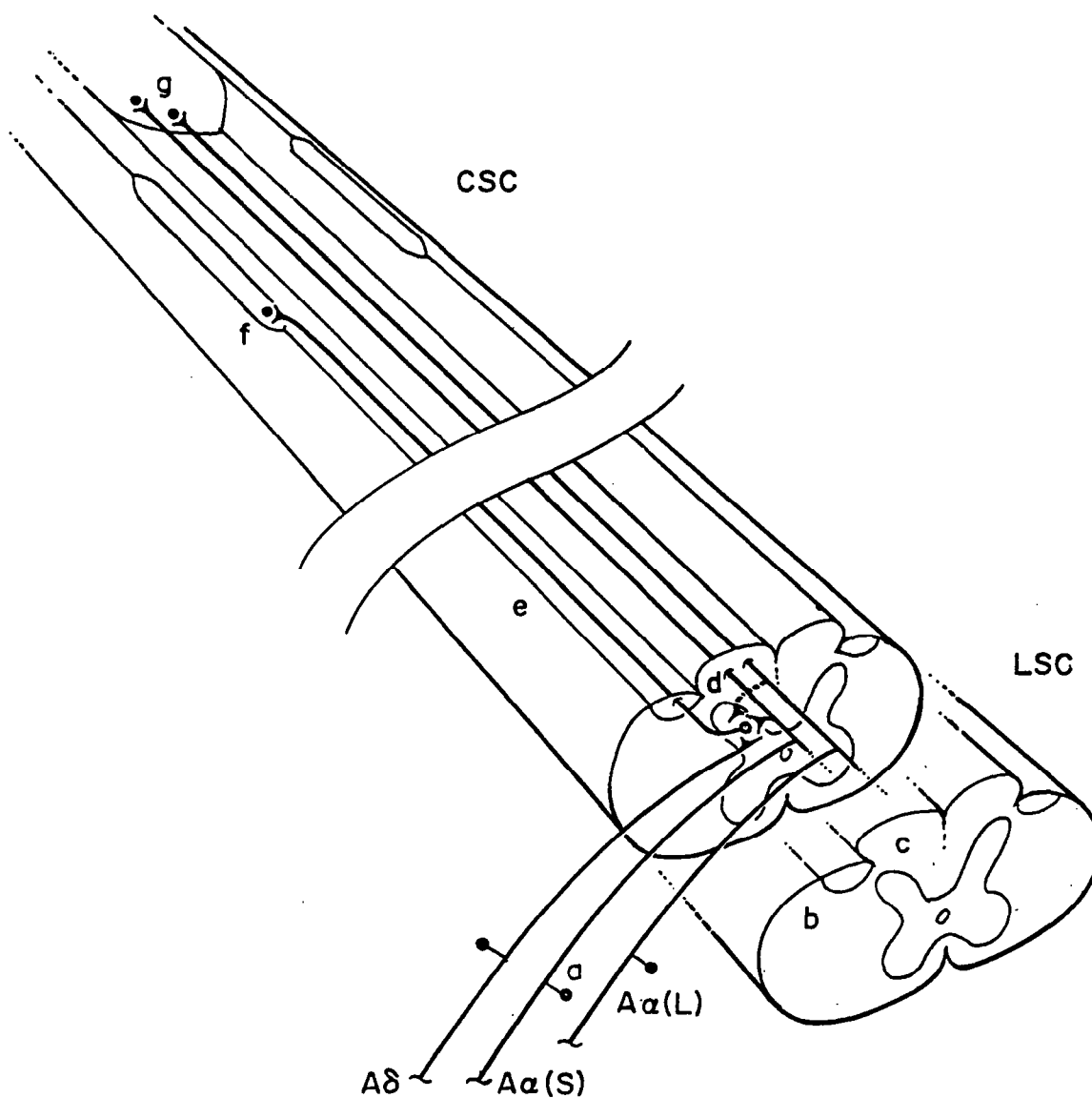


Figure 20. Projection of peripheral fibers onto spinal cord and lateral cervical nucleus. a, peripheral neural input; b, dorsolateral columns; c, dorsal columns; d, synaptic connections with neuron in lamina IV; e, spinocervical tract; f, lateral cervical nucleus; g, dorsal column nuclei; LSC, lumbar spinal cord; CSC, cervical spinal cord; A δ , small myelinated fibers of peripheral nerve; A α (S), A α (L), smaller and larger divisions of the larger myelinated fiber group of peripheral nerve. (110)

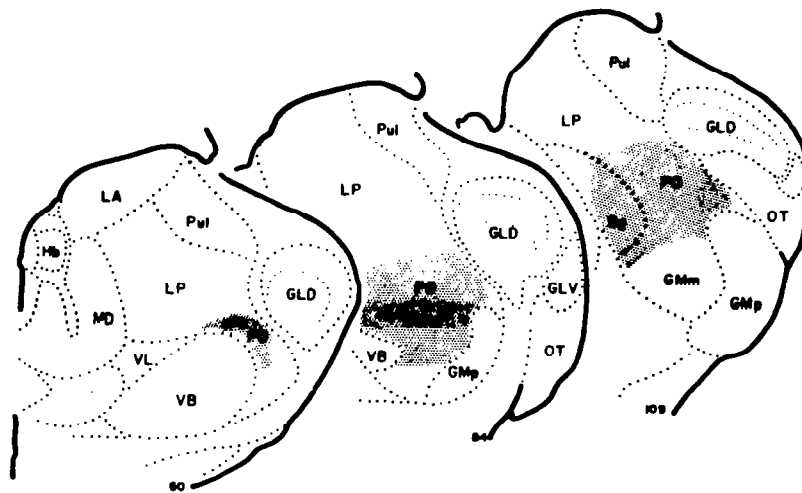


Fig. 21. Drawings of coronal sections through the posterior third of the dorsal thalamus of the cat. The sections were some 600 microns apart in the orocaudal dimension. The shaded region indicates the position of the posterior group of nuclei of the thalamus in relation to the ventrobasal complex. Abbreviations: GLD, dorsal nucleus of the lateral geniculate body; GLV, ventral nucleus of the lateral geniculate body; GMm, magnocellular division of the medial geniculate body; GMp, principal division of the medial geniculate body; Hb, habenular complex; LA, lateral anterior nucleus; LP, lateral posterior nucleus; MD, mediodorsal nucleus; OT, optic tract; PO, posterior group of nuclei of the thalamus; Pul, pulvinar; Sg, supragenicular nucleus; VB, ventrobasal nuclear complex of the thalamus; VL, ventrolateral nucleus. (88)

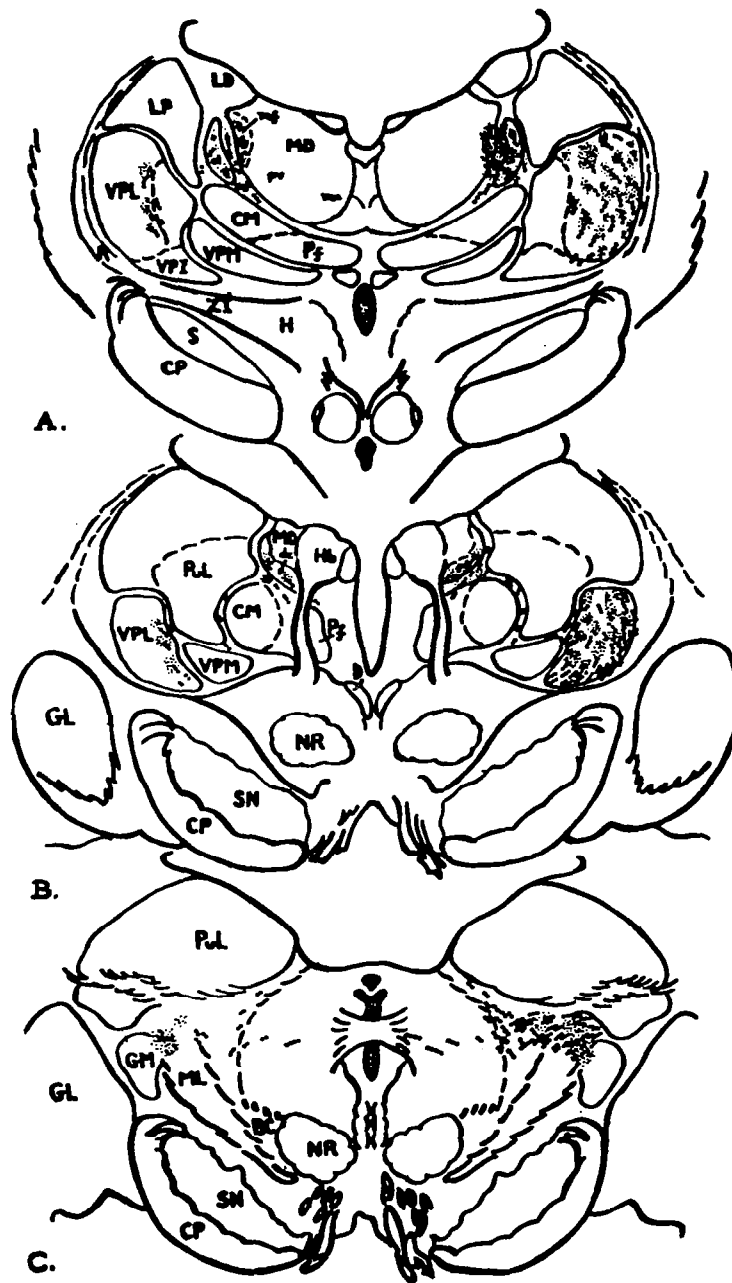


Figure 22. Distribution of terminal degeneration in the thalamus following anterolateral cordotomy. Cordotomy was performed on the left. Note cluster degeneration in the contralateral nucleus ventralis postero-lateralis and some clusters of degeneration in the medial portion of the ipsilateral nucleus ventralis postero-lateralis. (77)

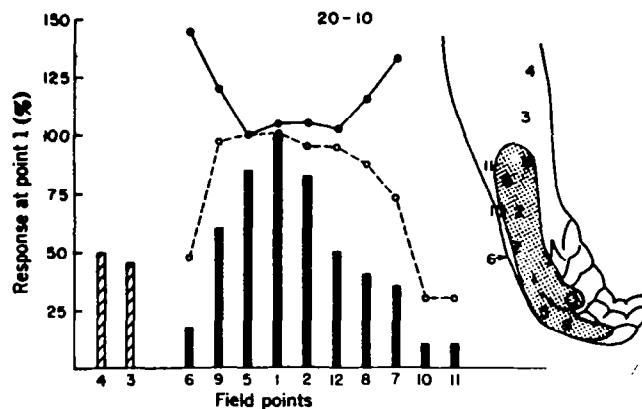


Figure 23. Gradations of the intensity of projection of the peripheral receptive field in the skin upon a cortical neuron of the postcentral gyrus of the macaque monkey. The bar graph indicates the mean value of the number of impulses evoked per electrical stimulus to the skin at each point marked on the drawing, plotted as percentage of the mean response of the point 1 population. Stimuli delivered at points 3 and 4, which are outside the field determined by physiological stimulation, were thought to activate nerve fibers innervating the field. Dashed lines plot the probability that a response will occur; solid lines, the mean latency of the population of responses, expressed as percentages of these values for the point 1 population. (88)

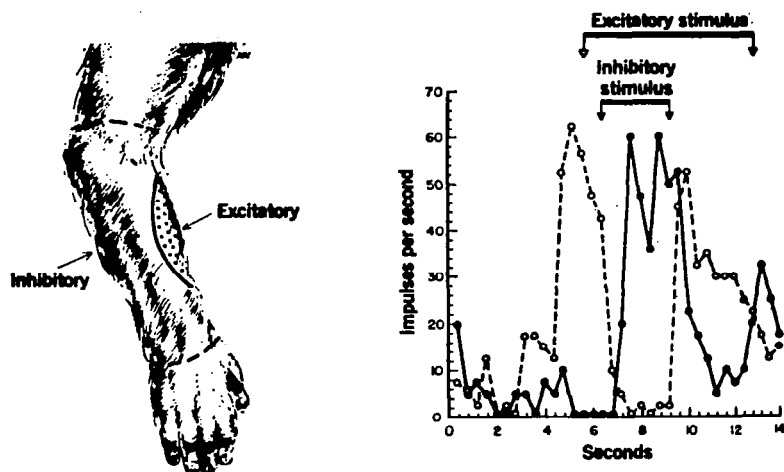


Figure 24. Afferent inhibition in the somatic system. A neuron of the post central gyrus of the macaque monkey was driven from the receptive field of the skin of the contralateral preaxial forearm, as shown in the drawing, and its discharge was inhibited by light mechanical stimulation within a much larger surrounding area, the inhibitory receptive field. This stimulation excited a second neuron whose discharges were also observed in the record, and the second neuron was inhibited by stimuli within the excitatory receptive field of the first. The reciprocal behavior of the two cells is indicated by the graph of impulse frequency versus time, the first cell by the dashed line, and the second by the solid line. (88)

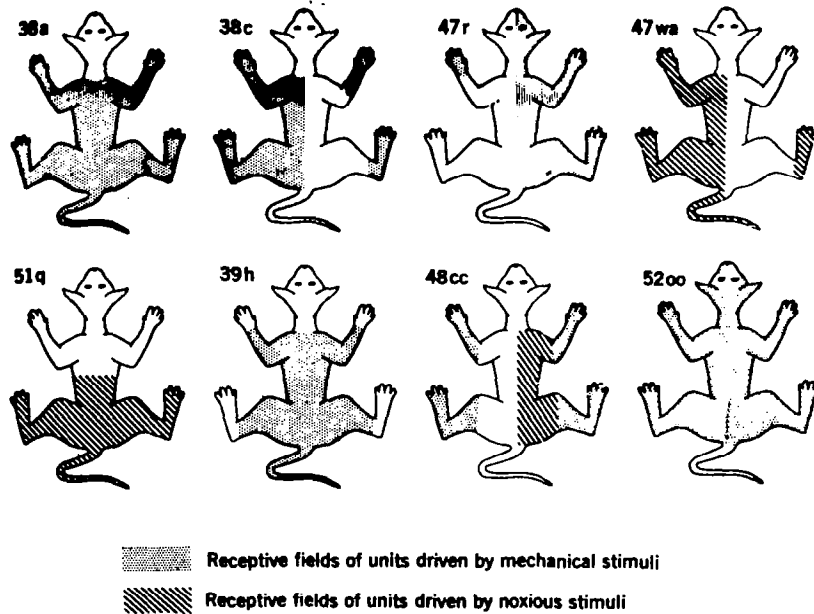


Figure 25. A representative sample of the peripheral receptive fields of neurons of the posterior nuclear group of the thalamus of the cat. Five of these cells (38a, 38c, 47r, 39h, and 52oo) were activated by light mechanical stimuli. Two were responsive to either form of stimulation, but the receptive fields for the two forms were not identical. The ipsilateral side of the body, relative to the recording electrode, is the right-hand side of each figurine drawing. (88)

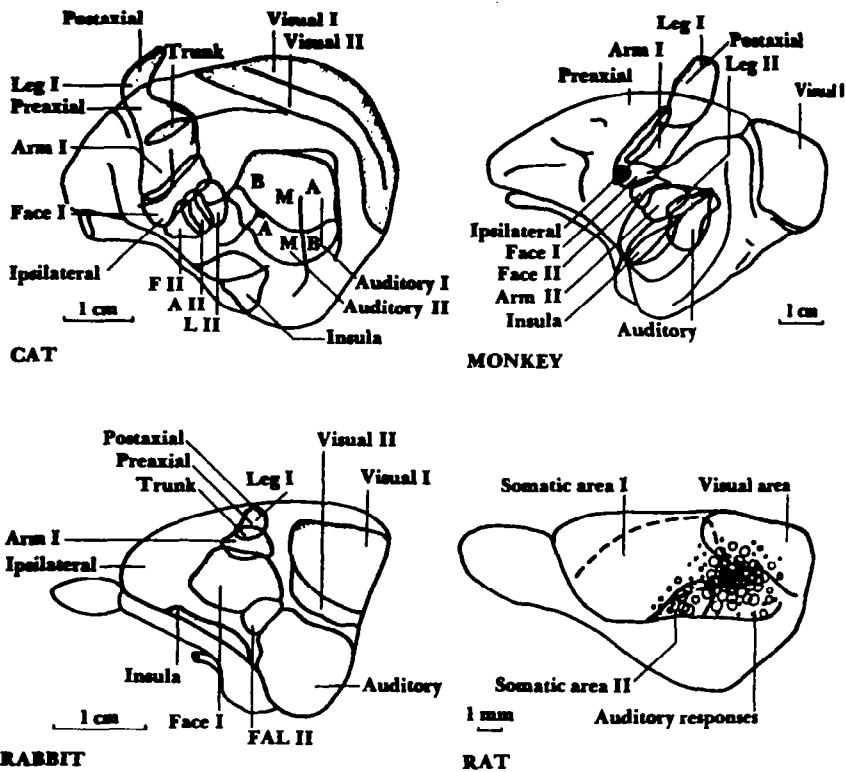


Figure 26. Somatic cortical areas of the cat, monkey, rabbit and rat in relation to the visual and auditory areas. (Adapted from C.N. Woolsey. Patterns of localization in sensory and motor areas of the cerebral cortex. In Milbank Memorial Fund. Twenty-seventh Annual Conference. The biology of mental health and disease. New York: Hoeber, 1952, p. 195.) (81)

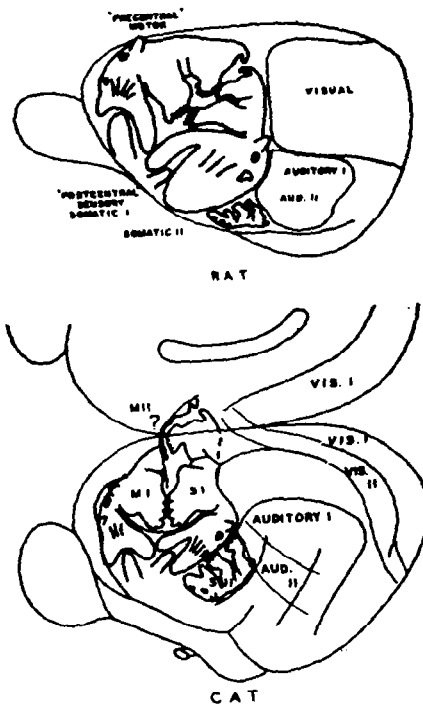


Figure 27. Above: diagram of cerebral cortex of rat showing general plan of organization of "postcentral sensory" and "precentral motor" areas. Below: Same for cat. MI = "precentral motor" and SI = "postcentral sensory" areas.

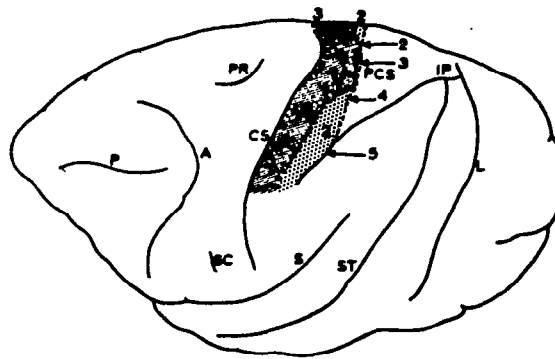


Figure 28. Reconstruction of the three cytoarchitectural areas of the somatic sensory cortex of the macaque monkey. Numbers without arrows indicate these areas. The junctions between them are considered to be regions of sharpest gradient in cellular changes, rather than sharp lines. Abbreviations as follows: A, arcuate sulcus; CS, central sulcus; IP, intraparietal sulcus; L, lunate sulcus; P, principal sulcus; PCS, postcentral sulcus; PR, superior precentral sulcus; S, sylvian fissure; SC, anterior subcentral sulcus; ST, superior temporal sulcus. (88)

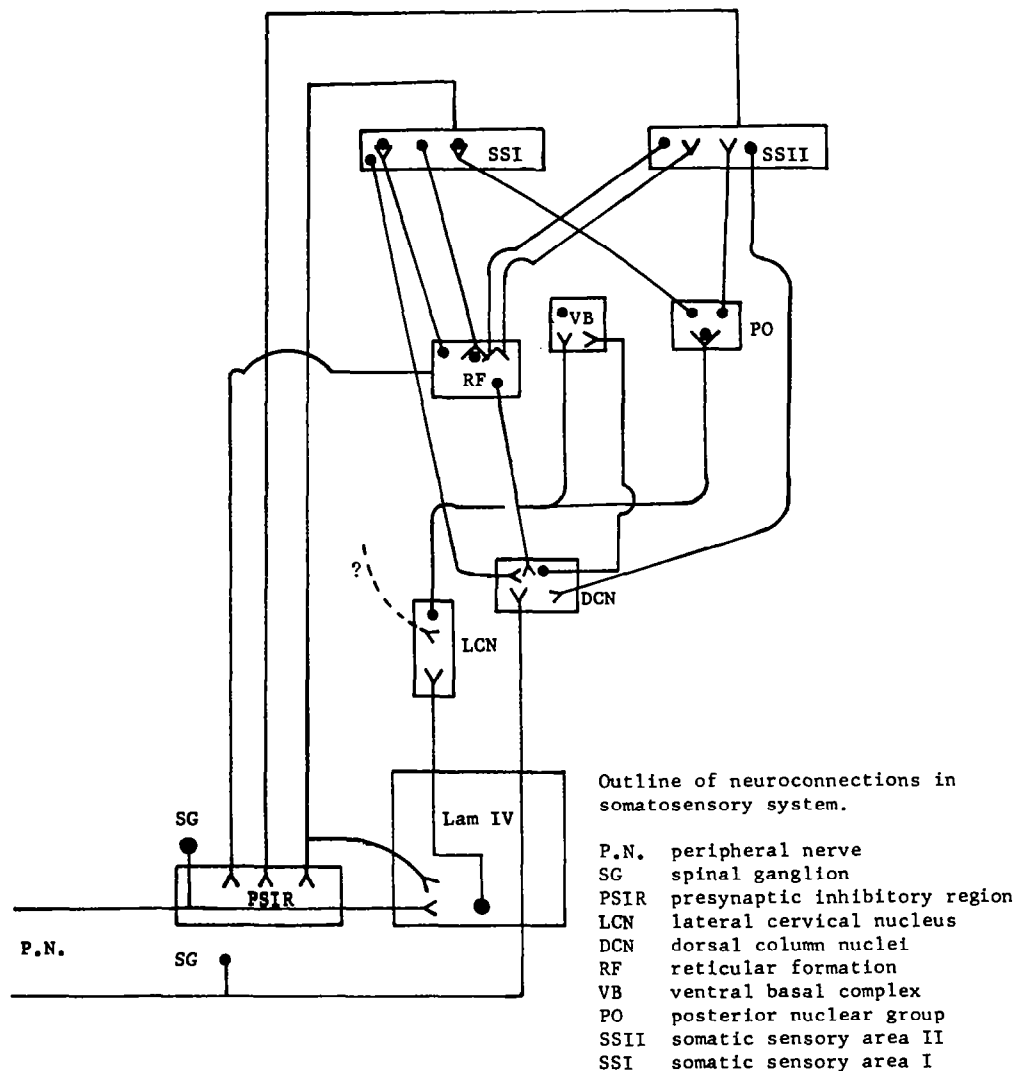


Figure 29. Overall scheme of the somesthetic system. Some interconnections of the somesthetic system. Two parallel pathways project to thalamic levels, one via the spinocervical tract, and the other via the dorsal columns. Both are inhibited by cortical and reticular mechanisms. Their modality characteristics and other properties are discussed in the text. The lateral cervical nucleus may be inhibited by higher central nervous system structures, but this has as yet not been demonstrated.

Bibliography

1. Adrian, E.D., M.C.K. Cattel and H. Hoagland. Sensory discharge in single cutaneous nerve fibers. *J. Physiol.*, 72:377-391, 1931.
2. Albe-Fessard, D. and A. Fessard. Thalamic integrations and their consequence at the telencephalic level. *Brain Mechanisms. In Progress in Brain Research*, Vol. I, Elsevier, Amsterdam, 1963. pp.115-154.
3. Amassian, V.E. Studies on organization of a somasthetic association area including a single unit analysis. *J. Neurophysiol.*, 17:39-54, 1954.
4. Andersen, P., Eccles, J.C., R.F. Schmidt and T. Yokota. Depolarization of presynaptic fibers in the cuneate nucleus. *J. Neurophysiol.*, 27:92-105, 1964.
5. Andersen, P., J.C. Eccles, R.F. Schmidt and T. Yokota. Slow potential waves produced in the cuneate nucleus by cutaneous volleys and by cortical stimulation. *J. Neurophysiol.*, 27:78-91, 1964.
6. Andersen, P., J.C. Eccles and T.A. Sears. Cortically evoked depolarization of primary afferent fibers in the spinal cord. *J. Neurophysiol.*, 27:63-77, 1964.
7. Andersson, S.A. Projection of different spinal pathways to the second somatic sensory area in cat. *Acta Physiol. Scand.* 56: Suppl. 194:1-74, 1962.
8. Armett, J.C., J.A.B. Gray and J.F. Palmer. A group of neurones in the dorsal horn associated with cutaneous mechanoreceptors. *J. Physiol. London* 156:611-622, 1961.
9. Barlow, H.B. The coding of sensory messages. In Current Problems in Animal Behavior, ed. by D.H. Thorpe and D.L. Zangwill. Cambridge U. Press, Cambridge, England, pp.331-336, 1961.
10. Békésy, G. von. Similarities between hearing and skin sensations. *Psychol. Rev.*, 66:1-22, 1959.
11. Bremer, F. Neurogenic factors influencing the evoked potentials of the cerebral cortex. In Sensory Communication, ed. W.A. Rosenblith, John Wiley & Sons: New York, 1961, p. 675-698.

12. Brodal, A. and B. Rexed. Spinal afferents to the lateral cervical nucleus in the cat. An experimental study. *J. Comp. Neurol.*, 98:179-211, 1953.
13. Busch, H.F.M. An anatomical analysis of the white matter in the brain stem of the cat. Van Gorcum, Assen., pp.116., 1961.
14. Carreras, M. and Andersson, S.A. Functional properties of neurons of the anterior extosylvian gyrus of the cat. *J. Neurophysiol.*, 26:100-126, 1963.
15. Chambers, W.W. and C.N. Liu. Cortico-spinal tract of the cat. An attempt to correlate the pattern of degeneration with deficits in reflex activity following neocortical lesions. *J. Comp. Neurol.*, 108:23-55, 1957.
16. Chang, H.T. The evoked potentials. Handbook of Physiology, Sec. I, Neurophysiology, Vol. I, pp.299-314, Am. Physiol. Soc., Washington, D.C., 1959.
17. Clark, D., J. Hughes and H.S. Gasser. Afferent function in the group of nerve fibers of slowest conduction velocity. *Am. J. Physiol.*, 114:69-76, 1935.
18. Collins, W.F. and F.E. Nulsen. Studies on sensation interpreted as pain, central nervous system pathway. *Clin. Neurosurg.*, 8:271-281, 1962.
19. Collins, W.F. and C.T. Randt. Fiber size and organization of afferent pathways. *Arch. Neurol.*, 5:202-209, 1961.
20. Cook, A.W. and Browder, E.J. Function of posterior columns in man. *Arch. Neurol.*, 12:72-79, 1965.
21. Dell, P.C. Humoral effects on the brain-stem reticular formation. In Reticular Formation of the Brain, ed. H.H. Jasper et al., Little-Brown: Boston, 1958, p. 365-380.
22. Dell, P.C. Reticular homeostasis and cortical reactivity. In Brain Mechanisms, Progress in Brain Research, Vol. I, Elsevier, Amsterdam, 1963. p. 82-114.
23. Douglas, W.W. and J.M. Ritchie. Mammalian non-myelinated nerve fibers. *Physiol. Revs.*, 42:297-334, 1962.
24. Dodt, E. Differential thermosensitivity of mammalian A fibers. *Acta Physiol. Scand.* 29:91-108, 1953.

25. Earle, K.M. Tract of Lissauer and its possible relation to the pain pathway. *J. Comp. Neurol.*, 96:96-111, 1952.
26. Eccles, J.C. Presynaptic inhibition in the spinal cord. In Progress in Brain Research ed. J.C. Eccles and J.B. Schade, Elsevier, Amsterdam, 1964, p. 73.
27. Erlanger, J. and H.S. Gasser. The action potential in fibers of slow conduction in spinal roots and somatic nerves. *Am. J. Physiol.*, 92:42-83, 1928.
28. Evans, M.H. Afferent fibers which mediate the reflex pupillary dilatation. *J. Physiol.*, 138:25-26P, 1957.
29. Evans, M.H. The spinal pathways of the myelinated and non-myelinated afferent nerve fibers that mediate reflex dilatation of the pupil. *J. Physiol.*, 158:560-572, 1961.
30. Fitzgerald, O. Discharges from the sensory organs of the cat's vibrissae and the modification in their activity by ions. *J. Physiol.*, 98:163-178, 1940.
31. Fjällbrandt, N. and I. Iggo. The effect of histamine 5-hydroxytryptamine and acetylcholine on cutaneous afferent fibers. *J. Physiol.*, 156: 578-590, 1961.
32. Frankenhaeuser, B. Impulses from a cutaneous receptor with slow adaptation and low mechanical threshold. *Acta Physiol. Scand.*, 18:68-74, 1949.
33. Gasser, H.S. Properties of the dorsal root unmyelinated fibers on the two sides of the ganglion. *J. Gen. Physiol.*, 38:709-728, 1955.
34. Gasser, H.S. Unmyelinated fibers originating in dorsal root ganglia. *J. Gen. Physiol.*, 33:651-690, 1950.
35. Geldard, S.A. Cutaneous channels of communication. In Sensory Communication, ed. W.A. Rosenblith, John Wiley & Sons Inc., New York, 1961. p. 73-88.
36. Gernandt, B.S. and Y. Zotterman. Intestinal pain: an electrophysiological investigation on mesenteric nerves. *Acta Physiol. Scand.*, 12:56-72, 1946.
37. Gordon, G. and M.G.M. Jukes. Correlation of different excitatory and inhibitory influences on cells in the nucleus gracilis of the cat. *Nature, Lond.*, 196:1183-1185, 1962.

38. Gordon, G. and M.G.M. Jukes. Cortical thresholds for excitation or inhibition of single cells in the nucleus gracilis of the cat. *J. Physiol.*, 169: 26-28P, 1963.
39. Gordon, G. and M.G.M. Jukes. Descending influences on the exteroceptive organizations of the cat's gracile nucleus. *J. Physiol.*, 173:291-319, 1964.
40. Gordon, G. and M.G.M. Jukes. Dual organization of the exteroceptive components of the cat's gracile nucleus. *J. Physiol.*, 173:263-290, 1964.
41. Gordon, G. and M.G.M. Jukes. An investigation of cells in the lateral cervical nucleus of the cat which respond to stimulation in the skin. *J. Physiol.*, 169:28-29P, 1963.
42. Gordon, G. and M.G.M. Jukes. Superimposition of sweeps on a storage oscilloscope as a method of studying changes in excitability of single nerve cells. *J. Physiol.*, 169:64-65P, 1963.
43. Gordon, G. and C.H. Paine. Functional organization in nucleus gracilis of the cat. *J. Physiol.*, 153: 331-349, 1960.
44. Gordon, G. and W.A. Seed. An investigation of nucleus gracilis of the cat by antidromic stimulation. *J. Physiol.*, 155:589-601, 1961.
45. Gray, J.A.B. Initiation of impulses at receptors. In *Neurophysiology*, Vol I of Handbook of Physiology, Am. Physiol. Soc., Washington, D.C. 1959, pp.123-146.
46. Gray, J.A.B. Sensory receptors. In Handbook of Physiology. Sec. I Neurophysiology Vol. I, pp. 123-146, Am. Physiol. Soc., Washington D.C., 1959.
47. Hagbarth, K.E. and J. Fex. Centrifugal influences on single unit activity in spinal sensory paths. *J. Neurophysiol.*, 22:321-338, 1959.
48. Hagbarth, K.E. and D.I.B. Kerr. Central influences on spinal afferent conduction. *J. Neurophysiol.*, 17: 295-307, 1954.
49. Holmqvist, B., A. Lundberg and O. Oscarsson. Supra-spinal inhibitory control of transmission to three ascending spinal pathways influenced by the flexion reflex afferents. *Arch. Ital. Biol.*, 98:60-80, 1960.

50. Hugelin, A. and M. Bonvallet. Etude experimental des anterieures lesiones reticulocortical . Proposition d'une theorie de l'asservissement reticulare a une systeme diffus corticale. J. Physiol. Paris 49:1201-1223, 1957.
51. Hunt, C.C. and A.K. McIntyre. An analysis of fibre diameter and receptor characteristics of myelinated cutaneous afferent fibers in cat. J. Physiol. London, 153:99-112, 1960.
52. Hunt, C.C. and A.K. McIntyre. Properties of cutaneous touch receptors in cat. J. Physiol. London, 153: 88-98, 1960.
53. Hunt, C.C. On the nature of vibration receptors in the hind limb of the cat. J. Physiol., 155:175-186, 1961.
54. Iggo, A. New specific sensory structures in hairy skin. Acta Neuroveg. 24:1-4, 1963.
55. Iggo, A. A single unit analysis of cutaneous receptors with C afferent fibers. In Ciba Foundation Study Group No. 1: Pain and Itch. Little Brown: Boston, 1959, pp. 41-59.
56. Jabbur, S.J. and A.L. Towe. Cortical excitation of neurons in dorsal column nuclei of cat, including an analysis of pathways. J. Neurophysiol. 24:499-509, 1961.
57. Kennard, M.C. The course of ascending fibers in the spinal cord of the cat essential to the recognition of painful stimuli. J. Comp. Neurol., 100:511-524, 1954.
58. Krauthamer, G. and Denise Albe-fessard. Inhibition of non-specific sensory activities following strio-pallidal and capsular stimulation. J. Neurophysiol., 28:100-124, 1965.
59. Kruger, L., R. Siminoff and I. Witkovsky. Single neuron analysis of the dorsal column nuclei and spinal nucleus of trigeminal in the cat. J. Neurophysiol., 24:333-349, 1961.
60. Kuypers, H.G.J.M., W.R. Fleming and J.W. Farinholt. Subcorticospinal projections in the Rhesus monkey. J. Comp. Neurol., 118:107-137, 1962.

61. Kuypers, H.G.J.M. and J.D. Tuerk. The distribution of the cortical fibers within the nuclei cuneatus and gracilis in the cat. *J. Anat. Long.*, 98:143-162, 1964.
62. Levitt, M., M. Carreras, W.W. Chambers and C.-N. Liu. Pyramidal influence on unit activity in posterior column nuclei of cat. *Physiologist* 3:103, 1960.
63. Lindblom, U.F. and J.D. Ottoson. Influence of pyramidal stimulation upon the relay of coarse cutaneous afferents in the dorsal horn. *Acta Physiol. Scand.*, 38:309-318, 1957.
64. Lloyd, D.P.C. The classification of galvanic reflex afferent fibers. *Proc. Nat. Acad. Sci.*, 48: 814-817, 1962.
65. Lloyd, D.P.C. The origins of dorsal root potentials. *J. Gen. Physiol.*, 32:409-443, 1949.
66. Lloyd, D.P.C. and A.K. McIntyre. Dorsal column conduction of group I muscle afferent impulses and their relay through Clark's column. *J. Neurophysiol.*, 13:39-54, 1950.
67. Lowenstein, W.R. On the "specificity" of a sensory receptor. *J. Neurophysiol.*, 24:150-158, 1961.
68. Lundberg, A. Supra-spinal control of transmission and reflex paths to motor neurons and primary afferents. In *Physiology of Spinal Neurons, Progress in Brain Research*, Vol. 12, ed. J.C. Eccles and J.B. Schade, Elsevier, Amsterdam, 1964, p. 201.
69. Lundberg, A. and O. Oscarsson. Three ascending spinal pathways in the dorsal part of the lateral funiculus. *Acta Physiol. Scand.*, 51:1-16, 1961.
70. McComas, A.J. Longitudinal organization in the gracile nucleus. *J. Physiol.*, 161:22-22P, 1962.
71. McComas, A.J. Responses of the rat dorsal column system to mechanical stimulation of the hind paw. *J. Physiol.*, 166:435-448, 1963.
72. McIntyre, A.K. and R.F. Mark. Synaptic linkage between afferent fibers of the cat's hind limb and ascending fibers in the dorsolateral funiculus. *J. Physiol.*, London, 153:306-330, 1960.

73. Magni, F., R. Melzack, G. Moruzzi and C.J. Smith. Direct pyramidal influences on the dorsal-column nuclei. Arch. Ital. Biol., 97:357-377, 1959.
74. Mann, S.J. and W.E. Straile. Tylotrich (hair) follicle: association with a slowly adapting tactile receptor in the cat. Science 147:1043-1045, 1965.
75. Manturana, H.R., J.Y. Lettvin, W.S. McCulloch, and W.H. Pitts. Anatomy and physiology of vision in the frog. J. Gen. Physiol., 43:129-175, 1960.
76. Mark, R.F. and J. Steiner. Cortical projection of impulses in myelinated cutaneous afferent nerve fibers of the cat. J. Physiol. London, 142:544-562, 1958.
77. Mehler, W.R., Feferman, M.E. and Nauta, W.J.H. Ascending axon degeneration following anterolateral cordotomy. Brain 83:718-750, 1960.
78. Melzack, R., G. Rose and D. McGinty. Skin sensitivity to thermal stimuli. Exp. Neurol., 6:300-314, 1962.
79. Melzack, R. and B. Schecter. Itch and vibration. Science, 147:1047-1048, 1965.
80. Mendel, L.M. and P.D. Wall. Presynaptic hyperpolarization --a role for fine afferent fibers. J. Physiol., 172: 274-294, 1964
81. Morgan, C.T. Physiological Psychology, McGraw-Hill: New York, 1965, 3rd ed.
82. Morillo, A. and D. Baylor. Electrophysiological investigation of lemniscal and paralemniscal input to the midbrain reticular formation. Electro-encephalog. Clin. Neurophysiol., 15:455-464, 1963.
83. Morin, F. Afferent projections to the midbrain tegmentum and their spinal course. Am. J. Physiol., 172: 483, 1953.
84. Morin, F. A new spinal pathway for cutaneous impulses. AM. J. Physiol. 183:245-252, 1955.
85. Morin, F. and L.M. Thomas. Spinothalamic fibers and tactile pathways in the cat. Anat. Rec., 121: 344, 1955.
86. Morin, F. and J.V. Catalano. Central connections of a cervical nucleus (nucleus cervicalis lateralis of the cat) J. Comp. Neurol., 103:17-32, 1955.

87. Morin, F., S.T. Kitai, H. Portnoy and C. Demirjian. Afferent projections to the lateral cervical nucleus: a microelectrode study. *Am. J. Physiol.*, 204:667-672, 1963.
88. Mountcastle, V.B. Some functional properties of the somatic afferent system. In Sensory Communication, ed. W.A. Rosenblith, John Wiley & Sons Inc.:403-436 New York, 1961.
89. Mountcastle, V.B., G.F. Poggio and G. Werner. The relation of thalamic cell response to peripheral stimuli varied over an intensive continuum. *J. Neurophysiol.*, 26:807-833, 1963.
90. Mountcastle, V.B. and T.P.S. Powell. Neural mechanisms subserving cutaneous sensibility, with special reference to the role of afferent inhibition in sensory perception and discrimination. *Bull. Johns Hopkins Hosp.* 105:201-232, 1959.
91. NorrSELL, U., and P. Voerhoeve. Tactile pathways from the hind limb to the cerebral cortex in cat. *Acta Physiol. Scand.* 54:9-17, 1962.
92. Nyberg-Hansen, R. and A. Brcdal. Sites of termination of corticospinal fibers in the cat. An experimental study with silver impregnation methods. *J. Comp. Neurol.* 120:369-392, 1963.
93. Oswaldo-Cruz, E., and C. Kidd. Functional properties of neurons in the lateral cervical nucleus of the cat. *J. Neurophysiol.* 27:1-14, 1964.
94. Perl, E.R., D.G. Whitlock and J.R. Gentry. Cutaneous projection to second-order neurons of the dorsal column system. *J. Neurophysiol.* 25:337-358, 1962.
95. Pillsbury, D.M. and W.B. Shelley. The skin and its appendages. In Principles of Internal Medicine by Harrison et al., Blakeston: New York, 1962, p. 71-87.
96. Poggio, G.F. and Mountcastle, V.B. A study of the functional contributions of the lemniscal and spinothalamic systems to somatic sensibility. *Johns Hopkins Hosp. Bull.*, 196:266-316, 1960
97. Ranson, S.W. The course within the spinal cord of the non-medullated fibers of the dorsal root; a study of Lissauer's tract in the cat. *J. Comp. Neurol.*, 23:259-281, 1913.

98. Ranson, S.W. and P.R. Billingsley. The conduction of pain for afferent impulses in the spinal nerves. Studies in vaso-motor reflex arc -II. AM. J. Physiol., 40: 571-584, 1916.
99. Rexed, B. A cytoarchitectonic atlas of the spinal cord in the cat. J. Comp. Neurol., 100:297-379, 1954.
100. Rexed, B. Some aspects of the cytoarchitectonics and synaptology of the spinal cord. In Organization of the spinal cord, Progress in Brain Research, Vol. II. ed. J.C. Eccles and J.B. Schade, Elsevier, Amsterdam, 1964, p. 65.
101. Rexed, B. and G. Strom. Afferent nervous connections of the lateral cervical nucleus. Acta Physiol. Scand. 25: 219-229, 1952.
102. Rosner, B.S. Neural factors limiting cutaneous spacial temporal discriminations. In Sensory Communication, ed. W.A. Rosenblith, John Wiley & Sons Inc.: New York, 1961, p.725-738.
103. Scherrer, H. and Hernandez-Peon, R. Inhibitory influence of reticular formation upon synaptic transmission in gracilis nucleus. Fed. Proc. 14:132, 1955.
104. Sprague, J.N. and H. Ha. The terminal fields of the dorsal root fibers in the lumbo-sacral spinal cord of the cat and the dendritic organization of the motor nuclei. In Organization of the Spinal Cord, Progress in Brain Research, Vol. 11, ed. J.C. Eccles and J.B. Schade, Elsevier: Amsterdam, 1964, p. 133.
105. Stevens, S.S. The psychophysics of sensory function. In Sensory Communication, ed. W.A. Rosenblith. John Wiley & Sons Inc: New York, 1961, p. 1-35.
106. Straile, W.E. Sensory hair follicles in mammalian skin, the tylotrich follicle. AM. J. Anat., 106:133-148, 1960.
107. Tapper, D.N. Cutaneous slowly adapting mechano-receptors in the cat. Science 143:53-54, 1964.
108. Taub, A. Local, segmental and supraspinal interaction with the dorsolateral spinal cutaneous afferent system. Exp. Neurol., 10:357-374, 1964.
109. Taub, A. Unpublished observations, 1965.
110. Taub, A. and P.O. Bishop. The spinocervical tract, dorsal column linkage, conduction velocity, primary afferent spectrum, in press, 1965.

111. Toennies, J.F. Reflex discharge from the spinal cord over the dorsal roots. *J. Neurophysiol.*, 1:378-390, 1938.
112. Towe, A.L. and S.J. Jabbur. Cortical inhibition of neurons in dorsal column nuclei of cat. *J. Neurophysiol.*, 24:488-498, 1961.
113. Towe, A.L. and I.D. Zimmerman. Peripherally evoked cortical reflex in the cuneate nucleus. *Nature*, Lond., 194:1250-1251, 1962.
114. Wall, P.D. Cord cells responding to touch, damage and temperature of skin. *J. Neurophysiol.*, 23:197-210, 1960.
115. Wall, P.D. Presynaptic control of impulses at the first central synapse in the cutaneous pathway. In *Physiology of Spinal Neurons, Progress in Brain Research*, Vol. 12, ed. J.C. Eccles and J.B. Schade, Elsevier, Amsterdam, 1964, p. 93.
116. Wall, P.D. Two transmission systems for skin sensation. IN *Sensory Communication*, Ed. W.A. Rosenblith, Wiley & Sons Inc: New York, 1961, pp. 475-496.
117. Wall, P.S. and Cronly-Dillon, J.R. Pain, itch and vibration. *AMA Arch. Neurol.*, 2:365-375, 1960.
118. Wall, P.D. and A. Taub. Four aspects of trigeminal nucleus and a paradox. *J. Neurophysiol.*, 25:110-126, 1962.
119. Weddell, A.G.N. Normal and abnormal sensory patterns: Pain II. "Activity patterns" hypothesis for sensation of pain. In *Neurophysiopathology: Some Relationships of Normal to Altered Nervous System Activity, Progress in Neurobiology*, Vol. 5, ed. R.G. Grennell, Hober: New York, 1962, pp. 134-177.
120. Weddell, G., E. Palmer and D. Taylor. The significance of the peripheral anatomical arrangements of the nerves which serve pain and itch. In *Ciba Foundation Study Group No. 1: Pain and Itch; Nervous Mechanisms*. Little-Brown: Boston, 1959, pp. 3-12.
121. Winkelmann, R.K. *Nerve Endings in Normal and Pathologic Skin*. C.C. Thomas: Springfield, 1960.
122. Winkelmann, R.K. The sensory end organ of the hairless skin of the cat. *J. Inves. Dermatology* 29:347-352, 1957.

123. Zotterman, Y. The peripheral nervous mechanism of pain; a brief review. IN Ciba Foundation Study Group No. 1, Pain and Itch. Little - Brown: Boston, 1959, p. 13-25.
124. Zuben, J. Dermo-optical perception, a cautionary report. Science 147:985, 1965.

SECTION 6

THE CARDIOVASCULAR SYSTEM

by

Peter G. Katona

SUMMARY

This report on the control of the cardiovascular system is divided into two parts. The first part briefly describes quantitative aspects of the physiology of the control system, while the second part contains a more detailed review of attempts to characterize this system in engineering terms. The emphasis is on the discussion of the present understanding of blood pressure regulation by the central nervous system. Hormonal effects, blood volume regulation by the kidneys, the regulation of respiration and the chemical contents of the blood are not included in any detail in the report. The citation of literature is illustrative rather than encyclopedic.

PART I - PHYSIOLOGY

Introduction

Ever since Harvey discovered the circulation of blood in the 17th century, the action of the heart has been studied by physiologists with great interest. Many of the early theories dealing with the circulation were based on experiments performed on heart-lung preparations in which the heart was not experiencing central neural control. It was not until well into the present century that the importance of the central neural control of the circulation became fully understood. Then it was found that the circulatory system of the intact animals was under so many controls, that they often overshadowed the regulating mechanisms suggested by the heart-lung preparation experiments.

At the present still much work is being done to determine the various factors which control and the way they control the performance of the circulatory system. There have been many experimental results, some of them consistent, but some of them appear to be contradictory. It seems that real understanding of the system is no longer possible just by accumulating more and more data, but that there is a pressing need for mathematical formulation of the relationships in the system.

The first part of this report describes those functional relationships which are most important for the understanding of attempts to characterize the system in engineering terms. It stresses the effect and cause relationships, while omitting detailed anatomical considerations.

Definition of Variables and Basic Relationships

The heart and circulatory mechanism is a complex system which supplies the body tissues with nutrients, oxygen and other chemicals, and which also removes carbon dioxide and other waste products from the same tissues.

The circulation of blood is caused by the pressure difference developed by the pumping action of the heart. The arterial pressure is determined by three factors: the heart rate, stroke volume, and peripheral (or vascular) resistance. The heart rate is the number of heart beats in one minute, the stroke volume is defined as the volume of ejected blood for one heart beat, and the peripheral resistance is taken as a measure of the state of contraction of the peripheral blood vessels. Normally peripheral resistance is measured as the ratio of arterial pressure to the minute volume of blood flow. The latter quantity by definition is the product of stroke volume and heart rate, and is often referred to as cardiac output.

It is important to note that the usual definition of peripheral resistance is meaningful only when average pressure is divided by average flow, although even this quantity is a function of pressure. When dealing with constantly pulsating pressure and flow, dynamic effects (inertia, elasticity) cause a phase shift between pressure and flow. In these cases it is possible to define a complex "incremental impedance" which is analogous to the impedance commonly used in electric circuits. Normally, however, peripheral resistance is computed as a real number and is used only as a rough indication of the tone (state of contraction) of the blood vessels.

Heart rate, stroke volume, and peripheral resistance are quantities that are under direct neural control, and therefore they shall be considered as "primary variables". Blood pressure and cardiac output are "secondary variables" since they can be changed only by changing one or more of the primary variables.

The next section reviews the ways in which the primary variables are controlled.

Control of Primary Variables

1. The heart rate is under constant neural control of the brain, although the heart beats rhythmically after its complete separation from the central nervous system. The controlling nerves are the vagus (parasympathetic) and the cardio-

accelerator (sympathetic) nerves. The vagus, which is the dominant heart rate controller, has a constant braking effect on the heart: when it is cut the heart accelerates, and when it is stimulated the heart slows down. The cardiac accelerator nerve has an opposite effect, but often it appears to play a relatively minor role as long as the vagus is intact.

Some of the factors that influence heart rate are briefly summarized:

a) Blood pressure reflexly influences heart rate. When the pressure increases, the heart rate decreases, thereby counteracting the rise in blood pressure. When the pressure falls, the heart rate **rises**.

b) According to classical views, when the **pressure** in the right atrium increases ("increased venous return") the heart accelerates. This is known as the Bainbridge reflex, and it was assumed to play an important role in the control of circulation. More recently, however, considerable doubt has been cast on the importance of this reflex (34). Some researchers have denied (3) while others have supported its existence (17). It has also been shown that increased venous return may even result in a slowing of the heart when the initial heart rate is high, while acceleration usually occurs when the initial heart rate is low (8).

c) When body needs increase, as during exercise, the heart rate of untrained subjects goes up, often before the onset of the expected exercise (32).

d) Respiration effects heart rate, probably through stretch receptors in the chest. Both inspiration and expiration cause acceleration first and deceleration later. The analysis of this mechanism is an example showing how engineering approach can solve a physiological problem (5).

e) The chemical contents of the blood can influence heart rate both directly and through the chemoreceptors. The direct action of chemicals on the heart, such as cardiac slowing due to the CO_2 , is probably not part of a regular controlling mechanism. The chemoreceptors, which are sensitive to the O_2 and CO_2 content of the blood, reflexly influence heart rate in such a way as to cause cardiac acceleration in case of lack of oxygen.

2. The stroke volume is the difference between the initial and final volume of the left ventricle in the systolic phase of the cardiac cycle, and therefore it is dependent on the mechanical properties of the heart muscles.

According to the classical Starling's Law, the contractility of the heart is enhanced as increased inflow of blood increases the fiber length of the cardiac muscles. This implies that these fibers automatically gain the necessary energy to develop, within a certain limit, whatever stroke volume is necessary to deliver all the blood flowing into the heart.

In addition to initial fiber length, sympathetic nerve impulses from the brain also influence the contractility of

the heart. This influence seems to play an important role in controlling the circulation, since it has been found that increased stroke volume may be delivered when the heart is contracted (33). Parasympathetic effects are believed to have little influence on the contractility of the heart. Chemical substances directly effect the heart muscles in a number of ways, but under normal conditions they are not assumed to have constant controlling functions.

It is known that in trained subjects, like athletes, an increased stroke volume is an important factor in providing increased blood flow during exercise. For these subjects the increased stroke volume is often not accompanied by an increased heart rate, while untrained subjects respond to exercise with a pronounced increase in heart rate and no change in stroke volume. The reason for the two types of responses to exercise is unknown.

A decrease in blood pressure has been shown to reflexly increase stroke volume at constant heart rate (33).

3. The vascular resistance is determined by the state of contraction of the smooth muscles that surround the blood vessels. The contraction of the muscles is under constant sympathetic neural control from the brain, so that the resistance of **many** and probably all segments of the circulation can be centrally controlled.

Chemicals have an important effect on the mechanical properties of blood vessels, and thus on vascular resistance. Norepinephrine, the substance that is liberated at the sympathetic nerve endings, is the most effective agent acting on the smooth muscles directly, but water and electrolyte concentrations, regulated by the kidney, are also instrumental in determining the properties of the vessel walls.

Vascular resistance is the most important single factor in determining blood pressure and it is reflexly influenced by the arterial pressure. Venous pressure has also been shown to effect vascular resistance (McDowell reflex), but some of the experimental results dealing with this effect are contradictory (3).

Regulating Systems

The previous section described how the primary variables of the circulation are controlled, but did not discuss how these controlling mechanisms are involved in the regulation of the secondary variables. This section deals with the regulation of the secondary variables.

1. The blood pressure regulatory system is the most studied cardiovascular regulatory mechanism. Long term regulation is provided by the kidneys, while a reflex initiated by the pressure receptors in the aortic arch and carotid sinuses compensates for sudden changes in arterial pressure.

By controlling the urinary output, the kidneys regulate the amount of water and electrolytes retained by the body. The retained water influences blood pressure through changing the blood volume. The water and electrolyte contents of the vessel walls also effect blood pressure through the modification of vascular resistance.

Renal activity is modified by renal arterial pressure, neural stimuli and hormonal mechanisms. Low renal arterial pressure decreases urinary output, and conversely, high renal arterial pressure increases urinary output. Increased sympathetic stimulation, which may be caused by a fall in systemic arterial pressure, causes the kidneys to release less fluid at a given renal pressure, thus helping to counteract the change in systemic pressure. The complex renal-aldosterone process, which is activated by the formation of renin as a result of a decrease in the renal perfusion pressure, also helps regulating the retention of water and electrolytes. The angiotensin that is generated in this process has a strong constricting effect on the smooth muscles of the vessel walls.

The regulation of pressure through the arterial pressure receptors is relatively well understood, and a simplified block diagram of the system is given in Fig. 1. The pressure receptors respond to changes of the stretch of the arterial wall by changing their neural firing frequency. The change in the neural activity is transmitted to the vasomotor area

of the brain. This area appears to be located in the medulla, but it may be under the influence of higher brain centers (26). The vasomotor area responds with changing the neural firing frequency on the efferent sympathetic and parasympathetic pathways. This change in neural activity influences heart rate, stroke volume and vascular resistance (including the capacity of venous vessels) in such a way as to bring the arterial pressure back to its original control level.

An important property of the pressure receptor reflex is that the receptors are sensitive not only to the general level but also to the time derivative of the pressure. This property and the nonlinear behavior of receptors are responsible for the observation that a pulsatile pressure waveform is a more effective input to the system than a steady pressure waveform of the same mean value (10).

One of the most important unsolved problems in the neural control of blood pressure is the question of "resetting" of the receptors. It has been observed that receptor activity decreases as a result of chronic high blood pressure (24), which reduces the effectiveness of the compensating action of the regulating mechanism. The exact cause of the altered activity of the receptors has not been established yet.

2. In addition to regulating blood pressure, there seems to be a mechanism that regulates cardiac output. It has already been mentioned that in untrained subjects increased

cardiac output during exercise is brought about by an increase in heart rate while the stroke volume remains constant. However, if during exercise the heart rate of unanesthetized dogs is artificially kept constant at different values, then the stroke volume becomes adjusted in such a way that the cardiac output stays constant regardless of the heart rate (42). Peripheral resistance has been suggested to play an important role in the control of cardiac output during exercise (44), but the feedback path was assumed to be through the pressure receptors.

In another experiment the arterial blood pressure was rhythmically changed by pumping blood in and out of the aorta (28). It was found that the stroke volume of the heart became so adjusted that the total blood flow, stroke volume plus pump volume, remained constant for any constant diastolic pressure regardless of the frequency of pumping. It has not been determined what mechanism is responsible for the regulation of blood flow.

Conclusion

It has been shown that the circulatory regulation is comprised of many overlapping regulatory mechanisms, most of which are not well understood. The interrelationship of these mechanisms can be understood only after the quantitative characterization of the individual servo loops.

PART II - ENGINEERING

Introduction

In the first part of this report the physiology of the cardiovascular system was briefly described. The description emphasized the functional relationships between the several variables that characterize the circulation, but ignored the quantitative characterization of these relationships. This part of the report reviews those publications that made an attempt to characterize the control system or a part of that system using quantitative techniques that are usually employed in the engineering sciences.

There has been no attempt to describe the entire cardiovascular control system in a quantitative manner. As was described in Part I, the complexity of the system is so enormous that not only is quantitative data lacking, but often the evidence for the very existence or non-existence of a control loop is contradictory. For example, some workers claim that increased venous inflow and/or pressure causes a reflex increase of heart rate (Bainbridge reflex) (17), while others completely deny the effect (3), or find that the direction of the resulting heart rate change depends on the initial heart rate (8).

The complexity of the system and the lack of understanding of even the functional relationships imposes serious difficulties in the concise and quantitative representation of the control mechanism. At the same time, however, it seems that the understanding of the complex behavior may come about through the quantitative description of the components of the system. The authors of the great majority of the articles that will be reviewed emphasize that their efforts are to be considered only as first attempts to shed some light on limited portions of the system of Fig. 1, which in itself is an extremely simplified diagram of the blood pressure regulatory system **alone** (27).

Most of the research dealing with the cardiovascular control mechanism can be divided into two categories. The first category comprises those works which concentrate mainly on the mechanical aspects of the circulation, while the works in the second category are mainly concerned with the regulation of some of the circulatory variables. The study of the mechanical aspects is necessary in order to understand the constraints between the circulatory variables that are due to the circulation being a closed flow-system, while the study of the regulation is indispensable for the understanding of how constraints are introduced by the central nervous system, which tends to keep some of the variables within narrow limits.

This review concentrates on the control of the circulation, but first it briefly deals with a few recent works attempting to characterize the mechanics that govern the flow of blood.

Mechanical Properties

Overall Characterization

The pressure-flow-volume characteristics of the circulation have been investigated and modeled in several recent works (2,9,12,41). Although the details of these models are different, the basic ideas behind them are very similar. All of them divide the circulatory system into several segments and describe the mechanical impedance of these segments by relatively simple equations. The equations are then combined to form a description of the entire system. Since the number of equations is generally over twenty, an analog computer is usually used for obtaining the solution.

As an illustration of these methods Beneken's work is reviewed in more detail (2). This work is the most recently published one and it also seems to have the most promise for incorporating the control features of the circulation at a later stage of development.

Beneken considers the circulation to consist of eight segments: intrathoracic systemic arteries including the

aorta, the extra-thoracic systemic arteries, the extra-thoracic systemic veins, the intrathoracic systemic veins, the right ventricle, the pulmonary arteries, the pulmonary veins, and the left ventricle. Each segment is characterized by three equations. These equations are:

1) the equation of continuity:

$$V(t) = V_0 + \int_0^t [F_{in}(\tau) - F_{out}(\tau)] d\tau$$

where V and V_0 are respectively the instantaneous and initial volume of the segment, and F_{in} and F_{out} are the inflow and outflow respectively;

2) the static pressure-volume relationship:

$$P = f(V)$$

where P is the (uniform) pressure within the segment;

3) the equation expressing the assumption that the outflow from a segment into another is proportional to the pressure difference between the two segments:

$$F_{out} = \frac{1}{R} (P_1 - P_2)$$

where P_1 and P_2 are the pressures in the original and the following segments respectively, and R is the resistance to flow between the segments. In case of the systemic arteries inertial effects may be included by subtracting a term,

proportional to $\frac{dF_{out}}{dt}$, from the right hand side of the above equation.

The pressure-volume relationship is chosen to be approximated by a piecewise-linear model:

$$P = \begin{cases} \frac{1}{C} (V - V_u) & \text{if } V > V_u \\ 0 & \text{if } V \leq V_u \end{cases}$$

where V_u is the effective unstressed volume and C is the compliance of the segment. Ventricular action is represented by the same equations, but the compliance of the ventricles is considered as a time-varying parameter. Thus a drop in the compliance at a constant volume is taken to account for the pressure rise in the ventricles. The effect of cardiac valves is represented, in effect, by ideal diodes in series with resistances.

The three equations for each of the eight segments are combined by observing that the outflow from one segment is equal to the inflow to the next one. Parameter values for the various resistances, compliances and unstressed volumes are compiled from the literature and the equations set up on an analog computer. The relationships between several of the parameters and variables are explored, and they are found to be in substantial agreement with available, although often only qualitative data.

As a refinement of the above description Beneken also considers a much more detailed model of the heart. This model takes into account the properties of the ventricular heart muscle by considering that the muscle consists of one contractile and two elastic elements. Information about the length-force relationship of these elements is collected from the literature and the force is related to the ventricular pressure through geometric considerations. A simple geometry is considered, but one that allows the representation of a limited interaction between right and left ventricles. Atrial effects are also included using atrial compliances as basic parameters.

The refined model of the circulation yields fifty-seven equations which are again solved by an analog computer. Experimentation with the model generally yields results that are in reasonable agreement with the limited data available on the actual human circulation.

The evaluation and use of Beneken's model is seriously handicapped by the fact that the model does not encompass any of the control features of the circulation. The exclusion of the mechanism that regulates blood pressure, for example, makes it virtually impossible to compare the performance of the model with that of the intact circulation as far as systemic pressure is concerned. Beneken indicates, however, that work for including control effects is in progress, and it

seems that his present model is a reasonable first step toward the more complete modeling of the mechanics of the circulatory system.

Similar considerations apply to the other published characteristics of the uncontrolled circulation, although it seems that Beneken's refined model is the one that can be most easily adapted to include controlling effects on the heart itself.

Partial Characterization

In addition to works which characterize the entire uncontrolled circulatory system, several attempts have been made to use engineering concepts and techniques for the description of individual parts of the circulatory system.

One of the most thoroughly described portions of the circulation is the arterial system (25). The pressure-flow relationship of the aorta has been discussed using the concept of complex impedance, and this impedance has been determined by power spectrum analysis (1). The results are difficult to incorporate in an overall view of the circulation and therefore are not reviewed here.

An interesting work by Roston (31) uses a two-chamber model to describe the blood pressure waveforms in both the ascending and descending aorta. Roston assumes that the aortic arch acts as a non-elastic passageway between the two elastic chambers representing the two parts of the aorta, and

that the cardiac output has the time dependence of a half-wave rectified sine-wave. The solution is obtained by Laplace transforms. It shows that the pressure waveforms in the two chambers can be substantially different from each other, in agreement with experimental observations. This result has clinical significance, because it shows that arterial pressure as measured in the brachial artery may not necessarily be taken to be representative of the pressure in the descending aorta.

Regulation of Blood Pressure

The engineering approaches to the description of blood pressure regulation will be discussed in three sections. The first section will deal with the characterization of some of the individual elements of the control system, the second section will deal with descriptions of parts of the system containing several elements, and the third section will deal with the characterization of the overall control properties of the regulatory mechanism.

Elements of the Control System

The pressure receptors. Landgren's classic work on the input-output characteristics of the pressure receptors (18) has been discussed from an engineering point of view by Pruslin (29). He presents a model that describes in a concise manner many of Landgren's observations. This model involves

three operations and is shown in Fig. 2. The first operation accounts for saturation effects, the second one for the rate-sensitivity of the receptors, and the third one for threshold effects.

This model qualitatively agrees with the following experimental observations:

a) The frequency of neural firing is dependent on the time derivative of the pressure waveform in addition to its steady state of value. This behavior is described by the linear transfer function,

$$H(s) = \frac{s + a}{s + b}, \text{ where } a < b$$

b) The response to a step increase in pressure is normally accomplished by an immediate rise in firing rate, which then gradually decays to a frequency which is higher than the original frequency prior to the pressure rise. The linear transfer function predicts this decay to be exponential, although Landgren found that a negative-power relationship well described the time course of the decay. A time constant of 0.1 sec is in reasonable agreement with the observed data.

c) A step decrease in pressure often causes a complete cessation of firing. Whether the nerves resume firing or not depends on the final steady state level of pressure. Fig. 3 illustrates how $H(s)$ followed by a threshold network can give rise to this behavior.

d) As the magnitude of a step rise in pressure is increased, the firing frequency tends toward a fixed response which still shows an initial overshoot. The observed responses can be accounted for by placing a saturating non-linearity before the linear transfer function, $H(s)$.

Pruslin made some experiments to substantiate his model. The results suggested the validity of the model, but the lack of accurate data prevented the drawing of a definite conclusion.

The relationship between blood pressure and firing rate was characterized by Warner (45) in a somewhat different manner. He found the following equation to hold above threshold:

$$n(t) = c_1 \left[(c_2 - \bar{p}) \frac{dp^+(t)}{dt} + (\bar{p} - c_3) \frac{dp^-(t)}{dt} + c_4 \frac{p(t) - (c_5 \bar{p} + c_6)}{c_7 + p(t) - (c_5 \bar{p} + c_6)} \right] \quad \text{for } n(t) > 0 \quad (1)$$

where $n(t)$ is the firing frequency, \bar{p} is the average pressure (as obtained by an RC integrator with a time constant of 7 sec), $p(t)$ is the instantaneous pressure, $\frac{dp^+(t)}{dt}$ and $\frac{dp^-(t)}{dt}$ are positive and negative pressure derivatives respectively, and c_1, c_2, \dots, c_7 are constants. The rate-sensitivity of the receptors is expressed by the first two terms of the equation, while saturation is taken into account by the last term. An overshoot followed by a gradual decay as a result of a step

change in pressure is also predicted by the last term. The time dependence of this term is caused by \bar{p} , which changes for 7 seconds as a linear function of time following a step change in pressure. The discussion on how well this model can describe the behavior of pressure receptors under a variety of conditions is lacking in Warner's paper.

It is obvious that Pruslin's and Warner's model predict very different responses to a step rise of pressure. Pruslin's model yields an instantaneously increased neural firing frequency which then decays to its steady state level with a time constant of 0.1 sec, essentially reaching that level in about 0.5 sec. Warner's equation yields infinite frequency at the instant of the pressure rise, after which the firing frequency decreases from a high but finite value to its steady level. The decrease lasts for seven seconds and it is linear with time if saturation is neglected.

It should be noted that Landgren originally fitted the relationship $f(t) = f_0 + at^{-b}$ to describe the firing frequency $f(t)$ after a step rise in pressure at $t = 0$, and this equation also gives an infinite firing frequency at $t = 0$. Thus it seems that Pruslin's model does not emphasize sufficiently the initial rise in pressure, while an infinite firing frequency for an infinitesimally short time seems to be unrealistic in Warner's model. It appears that the type of model Pruslin proposed may be improved by assuming two time

constants for $H(s)$: one would be very short and account for the large initial rise, the other one would be on the order of a few seconds and give rise to the adaptive behavior described by the last term in Warner's equation. This representation would combine the convenience of Fig. 2 with some of the accuracy afforded by the use of several parameters in the more cumbersome Eqn. 1. The data provided by Trank (38) may be used to check the validity of the above representation.

A completely different description of the pressure receptors was attempted by Teorell (37). He related blood pressure to the electrical response of the receptor nerve cells by differential equations and by assuming an N-shaped curve (analogous to the curve signifying a negative resistance in electronic circuits) between membrane resistance and "electro-osmotic flow". These equations defined a relaxation oscillator which could exhibit some of the characteristics of pressure receptor cells. These included threshold effects and the rate sensitivity of the response. The model, however, does not appear to be overly useful because it yields oscillations only at a particular frequency contrary to the behavior of the actual cells. This approach has not yet helped to describe the pressure receptor input-output characteristics which are of primary importance for the understanding of the regulatory mechanism.

The heart. The influence of sympathetic and parasympathetic vagal nerves on the heart rate and stroke volume are of great interest in the study of the blood pressure regulation. It appears, however, that studies using engineering-like approaches have dealt with the control of heart rate only.

Since the recording of normal activity from the cardiac vagus nerve involves great technical difficulties and has been successfully accomplished only very recently (13,46), most of the experiments dealing with the vagal control of heart rate involved the artificial electrical stimulation of the cut vagus nerve. Rosenblueth and Simeone (30) gave empirical equations of the steady state response of heart rate to electrical stimulation of the vagus and sympathetic nerves in 1934; in 1962 Warner and Cox (43) used step changes in uniform frequency of stimulation to derive nonlinear models of neural heart rate control. There seems to be no model that can describe the heart rate response to single vagal impulses first reported by Brown and Eccles (4).

The equations of Warner and Cox are based on the assumption that heart rate (or the reciprocal of heart rate, heart period, in the case of the vagus nerve) is proportional to the concentration of some chemical substance at the active site in the cardiac pacemaker. These equations express the

dynamics according to which the active chemical substances are assumed to be liberated by the neural impulses. The assumptions are not fully verified, but they are in agreement with the current very limited knowledge of the processes that occur at the sympathetic and parasympathetic nerve endings.

The models for sympathetic and parasympathetic control of heart rate result in a set of nonlinear equations which were simulated on an analog computer. The predicted output showed reasonable agreement with experimental data as long as sympathetic and vagal controls were considered separately, but superposition of the two controls did not hold. Katona (15) demonstrated that the nonlinear model of the vagal control of heart rate could be reduced to a simple RC integrator for weak and moderate vagal stimulation.

Van der Pol (40) discussed a model comprising three locked relaxation oscillators which could simulate some of the peculiar behavior of cardiac rhythm. He did not deal with the question of how neural impulses change heart rate.

Other components. There seems to be no significant work that has been successful in applying engineering techniques to the rest of the regulatory system of Fig. 1. The most interesting part of this system is the vasomotor area of the brain, but the present knowledge of its transfer characteristics is extremely sketchy and mainly qualitative. The only definite quantitative result seems to establish that the

transit time for the cardiovascular reflex is on the order of tens of milliseconds (18).

The control of vascular resistance by neural stimulation is also lacking quantitative description. Turner thoroughly reviewed the current views on this part of the control system, but his experimental data showed so much complexity and non-linearity that he failed to arrive at a concise model of the system (39).

Investigation of Portions of the Control System.

Heart rate is actively involved in the short-term regulation of blood pressure. Katona (14) investigated the relationship between these two variables considering arterial blood pressure as input and heart period as output. The experiments were conducted on anesthetized dogs, which had their circulatory mechanisms left intact. The blood pressure was changed by blowing up a balloon in the aorta of the experimental animals. The transient disturbances lasted 1 - 20 seconds.

It was found that the response of heart period to a change in blood pressure is highly nonlinear. The most conspicuous nonlinearity is that an increase in pressure generally results in a larger and faster increase in heart period than the decrease in heart period that is due to the same amount of decrease in pressure. As a first approximation, the system was assumed to be piecewise linear and thus characterized by

two impulse responses. One or the other impulse response was taken to represent the system depending on whether the pressure was above or below its steady state value. The impulse responses were determined by the numerical solution of the convolution equation that relates the input autocorrelation function to the input-output cross-correlation function. These functions were obtained by digital computer processing of pairs of input-output data involving pressure disturbances of arbitrary waveforms.

The computed impulse responses for the two regions were different, while those obtained from different pairs of records and representing the same region were quite similar, although dependent on the records that were used to obtain them. This indicated that the piecewise linear model could be considered only as a rough description of the system.

A recently developed improved model of the same system is shown in Fig. 4 (15,16). The input is a measure of the total neural activity of the pressure receptor nerves, and according to argument presented in the references it can be approximated by a linear combination of the systolic and diastolic blood pressures. The input activates one of two systems depending on whether it is larger or smaller than a certain threshold. This threshold often but not always is in the vicinity of the steady state value of the input. If the input is larger than the threshold (region A) the system is repre-

sented by an R-C-diode network with faster charging than discharging time. If the input is smaller than the threshold (region B) the system is represented by an overdamped second order transfer function. The sum of the responses of the two regions gives the total heart period response.

The above model was shown to describe the relationship between heart period and blood pressure for a variety of blood pressure disturbances, although some cases were also found where the model broke down. Operation in region A was attributed to the vagal control of heart period, while region B was associated with sympathetic effects.

An attempt by Lercari (20,21) to describe a portion of the pressure regulatory system comprising several of the building blocks in Fig. 1 involved the electrical stimulation of the cut pressure receptor nerves and measuring the corresponding changes of arterial pressure in the rabbit. A linear model was assumed to exist between the frequency of stimulation and the pressure, and the system was tested by applying sinusoidal variations in the frequency of the stimulating pulses. The resulting Bode plot indicated that a second order system with a double pole at 0.035 cps gave a reasonable approximation of the data, although a third order system with a break point frequency at 0.05 cps was preferable. The symmetrical response of the physiological system to positive and negative steps in the stimulating frequency was used as an evidence supporting the assumption of linearity.

There appears to be some contradiction between the results of Katona and Lercari. The system Katona investigated was found to be grossly nonlinear with the vasomotor area probably being responsible for much of the nonlinearity. Lercari found his system, which also involved the vasomotor area, to be essentially linear. The contradiction may arise from the possibility that Katona's description of the total number of neural firing on the carotid sinus nerve is unjustified and the nonlinearity he observed ~~was~~ mainly due to the carotid sinus mechanism. The contradiction may also be due to the artificial manner in which the carotid sinus nerves were stimulated in Lercari's experiments (he did not stimulate the C fibers), or may simply be attributed to species difference.

Control Properties of the Entire Blood Pressure Regulatory System

The studies that have been hitherto reviewed dealt with the description of the input-output characteristics of elements or groups of elements of Fig. 1, but they did not discuss how these characteristics determine the properties of the control system as a whole. The investigations to be reviewed next are all concerned with the description of the entire system from a control-theory point of view, and largely ignore the question of how the system components give rise to the control characteristics.

Warner investigated the frequency-dependent nature of blood pressure regulation in the dog by introducing sinusoid-

ally varying disturbances and measuring the blood pressure variations that these disturbances actually produced (41). The disturbances were introduced by stimulating the vagus nerve with an electrical impulse train the frequency of which was sinusoidally variable. It was found that the ability of the system to compensate for the disturbances was approximately constant below 0.033 cps and deteriorated at higher frequencies. This is shown by the dotted lines in the graph of Fig. 5.

Warner also examined the effect of increasing the "gain" of the feedback path of the control system by artificially amplifying the action of the carotid sinus by an electrical analog. This was achieved by generating an impulse train the frequency of which was proportional to the weighted sum of the arterial pressure and its first derivative. The impulse train was applied to the carotid sinus nerve in order to augment the natural activity on these nerves.

The results are illustrated by the solid lines in the graph of Fig. 5. They show that the pressure variations at most frequencies were larger when the analog was in the system than when only the normal regulatory mechanism of the dog was active. This indicates that an increase in the loop gain increased the steady state error for the control mechanism. This increase of steady-state error with increased loop gain was interpreted by Warner to mean that the increased loop gain

made the control system unstable, and the amplitude of the oscillations was limited only by the nonlinearities in the system. Guyton and Harris (11) also suggested that the instability of the control system may be responsible for the blood pressure oscillations often observed in dogs. These oscillations have a frequency on the order of 25 seconds, and they are most often observed when the condition of the dog deteriorates as a result of an excessive amount of blood loss.

It is difficult to draw quantitative interpretations of Warner's findings because of the indirect manner in which the disturbance was applied. It seems, however, that the sluggishness of the arterial smooth muscles has been established as an important factor in determining the behavior of the control system. From transient measurements the maximum response to pressure receptor stimulation was determined to be 10 to 15 seconds after the onset of the stimulation.

Stegemann (36) also investigated the frequency dependency of the blood pressure regulatory system of the dog, but he isolated the carotid sinus in order to perfuse it with an externally generated sinusoidal waveform. Thus his preparation operated in an open loop fashion except for the aortic nerves which he left intact. It was found that the average aortic blood pressure was not a pure sinusoid in contrast to the input pressure, but showed a relatively fast fall-time and slow rise-time. Periodic variations in arterial pressure were

demonstrated up to an input frequency of 1.3 cps, the highest investigated, showing that the controller was still effective at that frequency. It is quite apparent from the published records that regulation ~~was~~ taking place mainly by the fast vagal heart rate reflex.

In analyzing his data Stegeman plotted phase angles against frequency. These angles were measured between carotid sinus pressure maximum and aortic pressure minimum, and carotid sinus pressure minimum and aortic pressure maximum. They increase with frequency and show considerable scatter. Because of the difference between these two types of phase angles Stegeman concluded that two controllers are present, a sympathetic and a parasympathetic, and they each give rise to one of the phase angles. This conclusion is groundless **since** an apparent nonlinearity cannot be explained by the superposition of two linear systems which are implied by measuring phase angles.

A somewhat more systematic study of the same mechanism was made by Scher and Young (35). They also isolated and externally perfused the carotid sinus of both dogs and cats, but in many of their experiments they cut the vagus nerves in order to eliminate the effect of aortic receptors. It was found that the static incremental gain of the regulatory system shows a very definite peak at a certain input pressure. This pressure is in the vicinity of the normal blood pressure

of the laboratory animals. Step changes of input pressure are followed by a delay of approximately two seconds, after which the output slowly adjusts to its new level. The time required to reach 63% of the total change was found to be on the order of twenty seconds; sometimes an overshoot in the response was also observed.

When the input pressure was changed sinusoidally it was found that the output pressure also changed approximately sinusoidally. At low frequencies (below 0.01 cps) the input and output were exactly out of phase. For convenience this will be referred to as a phase angle of zero since the phase reversal is just due to the negative feedback nature of the system. Above 0.01 cps the phase angle started showing a lag which increased to 90° at 0.04 cps and to 180° at 0.07 to 0.10 cps. This phase difference is partially due to the pure delay in the system, and when that contribution ~~was~~ subtracted from the delay, the phase angle was never found to exceed 90° in the above range of frequencies. The magnitude of the gain of the system was found to be on the order of five at 0.01 cps which then decreased with increasing frequencies. In the vicinity of 0.1 cps the gain reduced to zero and the systemic blood pressure did not show periodic variations. It was still noted that the mean output pressure fell as the frequency of the input variations was increased above 0.1 cps. This is the rectifying property of the blood pressure reflex that had been studied previously (10).

The authors discussed the possibility of describing the reflex with equations based on their previous knowledge and experimental results, but they did not actually test any models. In effect they proposed a description similar to the one discussed by Pruslin (29), to account for the behavior of the pressure receptors. The efferent side of the reflex was proposed to consist of a pure delay followed by a first or second order dynamic system. The variation in steady state gain was to be taken into account by a power function.

The most comprehensive investigation of the overall properties of the blood pressure regulatory mechanism has been conducted by Levison (22). He perfused the carotid sinus of dogs with an externally generated pressure waveform which he could control by a servo-pump. For achieving a fully open-loop system without cutting the vagus nerves the aortic arch was denervated in several of the experiments. Because of the basically non-linear character of the mechanism a large variety of input waveforms were applied for testing the system. The results of some of these experiments are discussed next.

The static gain of the system was determined to be approximately two. When a high frequency (1 cps) sinusoidal waveform was superimposed on the input, the output pressure dropped, confirming the rectifying property of the system.

The sinusoidal response of the system was determined under two conditions. First, only the low frequency (less than

1 cps) disturbance was the input waveform; then a high frequency (more than 1 cps) variation was also added to the low frequency disturbance in order to simulate the effects of pulse pressure. The output in both cases was considered to be the low-frequency component of the arterial pressure. It was found that the gain curve for both cases reached a maximum between 0.03 and 0.05 cps and then started falling at a rate of 30 to 40 db/decade at higher frequencies. The addition of the high-frequency sine wave to the input pressure reduced the gain of the system by 20 to 50%, but left the phase relation unchanged. At very low frequencies (less than 0.01 cps) the input and output waveform were completely out of phase, which was defined as zero lag. The phase angle started lagging at increasing frequency, reaching 180° around 0.15 cps. At this frequency the gain was less than unity.

From the observation that the magnitude characteristics were similar to that of a second-order system but the phase lag continued beyond the additional 180° , Levison indicated the possibility of a 1.5 sec pure delay. This is in good agreement with the 2 second approximation of Scher and Young (34). The interpretation of Levison's frequency response data is difficult because of the distortion that the output sine wave often showed, but the data indicates the same trends that were found by Scher and Young.

An experiment clearly showing the nonlinear nature of the system consisted of applying a sinusoidally modulated (at

frequency. It appears that the pressure receptors should be characterized by a pole at a considerably higher frequency than 0.3 cps.

The great merit of Levison's work is sharply pointing out some of the nonlinear behavior of the pressure regulatory system. His model, however, should be considered only as a first step in the quantitative description of system characteristics.

Control of Cardiac Output

It was pointed out in the first part of this report, that during exercise an elevated cardiac output is maintained regardless of heart rate, which suggests the existence of a mechanism that regulates cardiac output. In a recent study, Topham (47) investigated the control of cardiac output during exercise by computer techniques.

The model that Topham used is shown in Fig. 7. This model is very similar to that of Fig. 1, except that it indicates that arterial pressure may have a direct effect on stroke volume, and it also considers local tissue environment as an essential factor in the regulatory system. Exercise is assumed to change the hypothetical "reference" setting in the brain, in addition to changing the local tissue environment due to the increased metabolism in the muscles. The tissue environments are, in turn, influenced by the amount of blood with

which they are supplied, which depends on the cardiac output.

The model was simulated on an analog computer. The various components of the model were characterized by previous knowledge (for example, the nonlinear models of Warner and Cox (43) were used for the determination of heart rate from sympathetic and vagal firing), or by simple relationships arrived at by reasoning on the basis of the physiology of the system. As an illustration, the vascular resistance was described by the following three equations:

$$R = R_0 + R_s - k_{16}s$$

$$\frac{dR_s}{dt} = k_{17}f_1 - k_{18}R_s$$

$$\frac{ds}{dt} = k_{14}(s_{\max} - s)(M_r + M_w) - k_{15}CO \cdot s$$

where

- R is the vascular resistance,
- R_0 is the resting value of resistance,
- R_s is the resistance due to sympathetic stimulation,
- s is the concentration of metabolite that causes vasodilatation,
- s_{\max} is the maximum of s that can exist,
- f_1 is the frequency of sympathetic nerve firing,
- M_r is the resting metabolism of metabolite,
- M_w is the metabolism during exercise,
- CO is the cardiac output,

and

$k_{14} \dots k_{18}$ are constants.

The results of the stimulation were shown only for one experiment. The predicted and experimental values of blood pressure, heart rate, stroke volume, resistance and cardiac output were in good agreement except for the falling blood pressure at the end of the exercise. Since there was no discussion on how well the model predicted the behavior of the physiological system under other conditions, and it is not surprising that over twenty parameters can be adjusted to fit five simple waveforms, the usefulness of the model cannot be assessed. The study seems to be quite incomplete without testing the model by keeping heart rate at various fixed values, and checking if stroke volume becomes adjusted in such a way that cardiac output remains unchanged.

Concluding Remarks

The foregoing sections of this report show that the application of engineering techniques to the analysis of blood pressure regulation has led to an improved understanding of the properties of the system but has not yet produced striking results. Ideally, a model of the physiological system should not only account for observed properties but should be able to predict the outcome of new experiments. It seems that at the present time the main problem is still fitting models to existing data, letting alone the question of prediction.

These investigations, however, provided extremely useful and numerous data on the various input-output characteristics

of the system. The data clearly indicates that nonlinearities play an important part in giving rise to these characteristics. The establishment of nonlinearity precludes the straight-forward application of the well-known methods for the analysis of linear control systems and this greatly contributes to the difficulty of analyzing the blood pressure regulatory mechanism.

It appears that in the future, effort should be expended in order to localize the physiologic origin of the nonlinearities that have been demonstrated to be present. In particular, it should be resolved to what extent the pressure receptors themselves were responsible for the nonlinearities in the regulatory system observed by Katona and Levison. An answer to this question should come from the quantitative determination of the input-output characteristics of the pressure receptors. Since this, just as other works aimed at determining the transfer properties of parts of the control system, requires the recording of neural impulses, a very close collaboration between physiologists and engineers is necessary.

The long-range success of understanding the control of circulation seems to hinge on the successful combination of the description of the mechanical (pressure-flow-volume) properties of the system with the description of how one variable influences the others through feedback loops via the central nervous system. Because of the extreme complexity, up to now the mechanical and control properties of the system have been

studied separately, which precludes the testing of derived models under normal physiological conditions. With the advance of computer technology it should be possible to simulate the entire control system in the not too distant future.

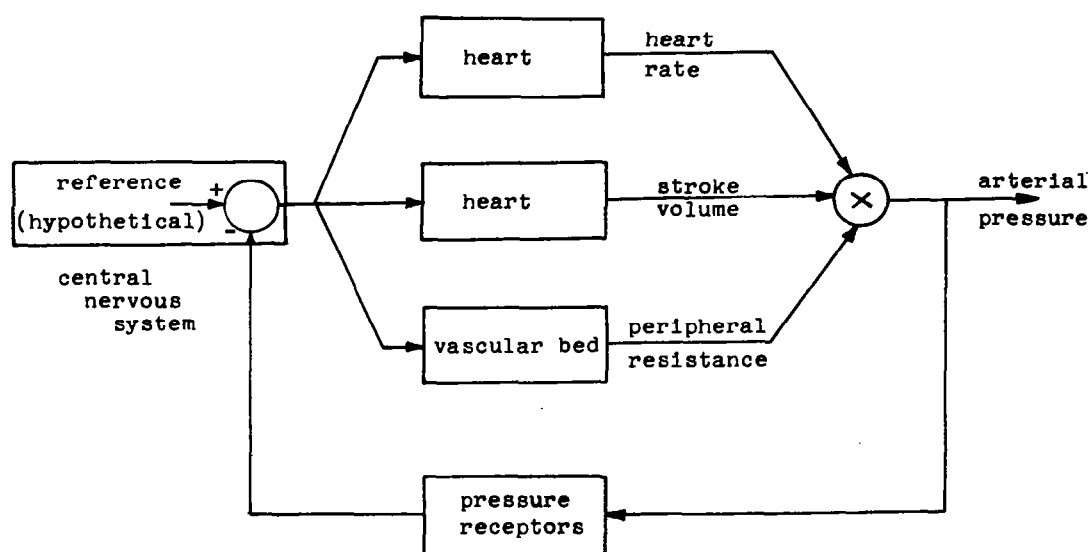


Figure 1. Pressure regulating system.

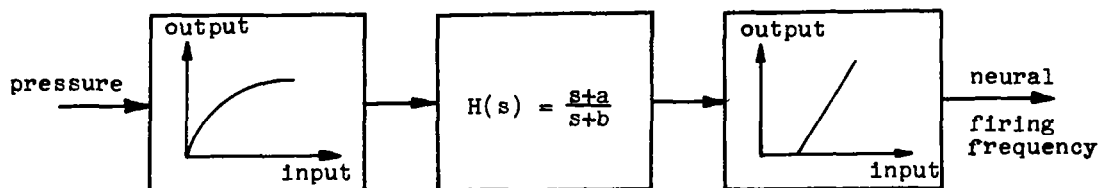


Figure 2. Model for the pressure receptors.

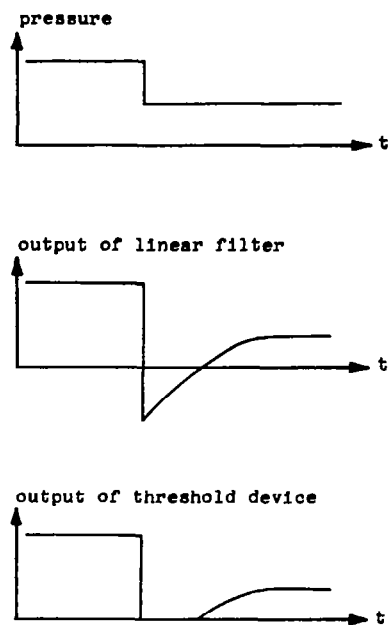


Figure 3. Simulation of pressure receptor nerve firing after a step decrease in pressure.

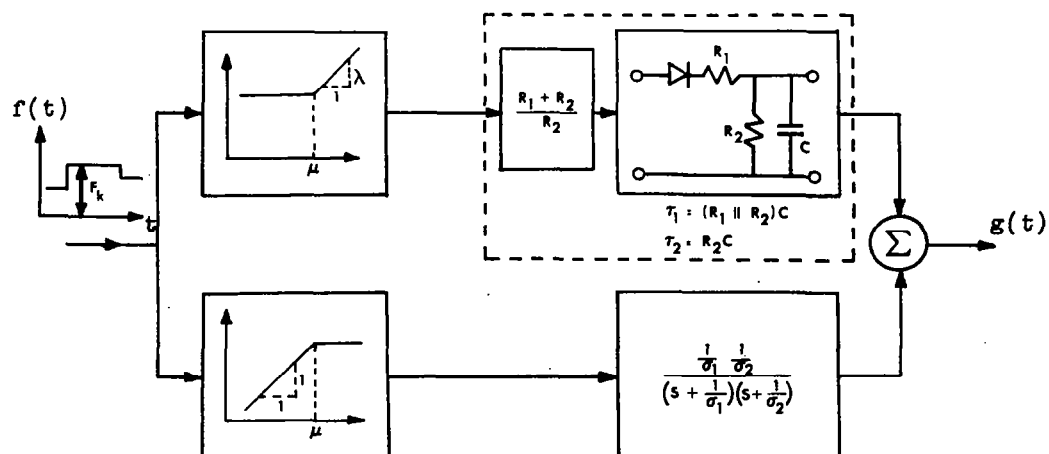


Figure 4. Model for the blood pressure control of heart period.

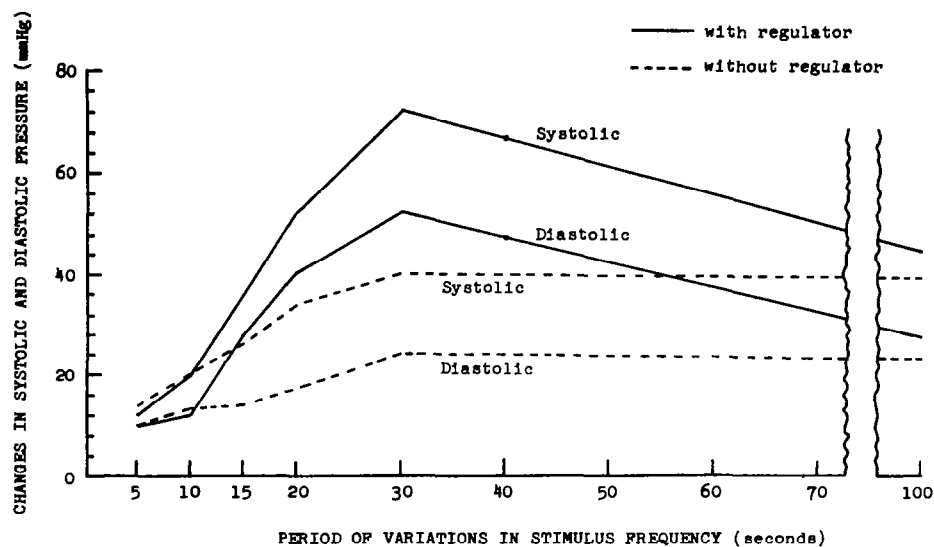


Figure 5. Changes in arterial pressure with and without amplification of the loop gain. (41)

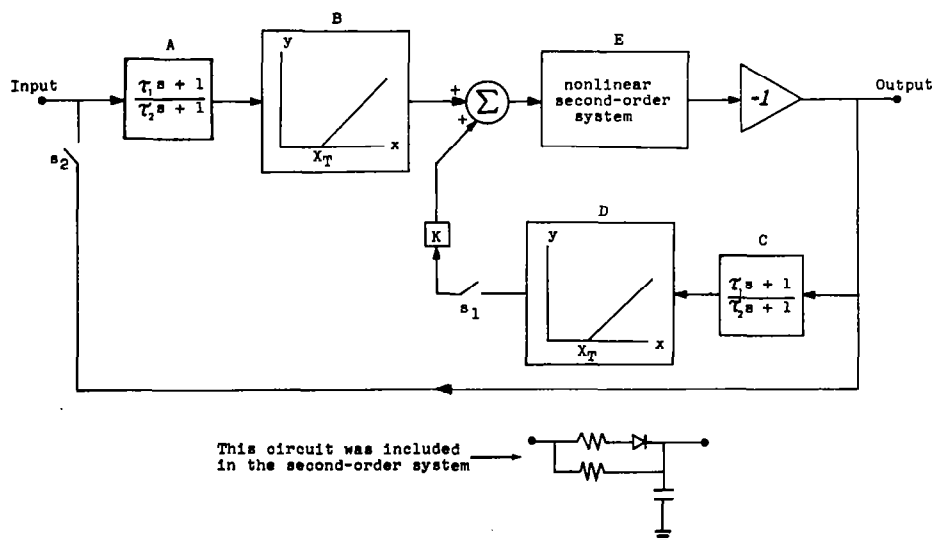


Figure 6. Model of the pressure receptor reflex system. (23)

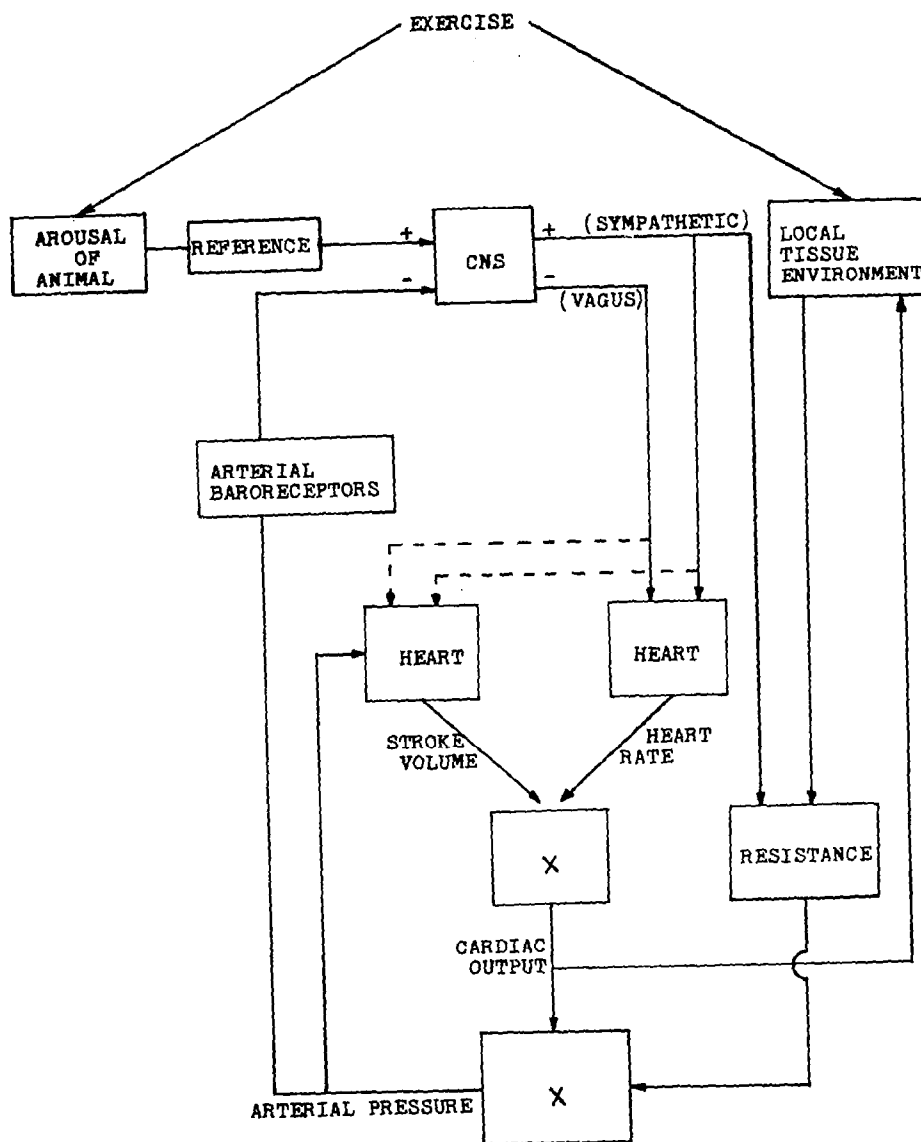


Figure 7. Model of cardiac control. (47)

BIBLIOGRAPHY

1. Barnett, G.O. and J.C. Michener. Power spectrum analysis of aortic impedance. Proc. of the 16th Annual Conf. on Engineering in Med. and Biol., Baltimore, 1963.
2. Beneken, J.E.W. A Mathematical Approach to Cardio-Vascular Function: The Uncontrolled Human System. Report No. 2-4-5/6, Institute of Medical Physics, Utrecht, Netherlands, 1965.
3. Blatteis, C.M. and S.M. Horvath. Inexcitability of the Bainbridge, McDowall and Harrison reflexes. Arch. Intern. Phys. Biochem., 72(4):535, 1964.
4. Brown, G.L. and J.C. Eccles. The action of a single vagal volley on the rhythm of the heart beat. J. Physiol., 82: 211, 1934.
5. Clynes, M. Respiratory sinus arrhythmia: laws derived from computer simulation. J. Appl. Physiol., 15(5):863, 1960.
6. Clynes, M. Unidirectional rate sensitivity: a biocybernetic law of reflex and humoral systems as physiologic channels of controls and communication. Annals of N.Y. Acad. of Sci., 92(3):946, 1961.
7. Clynes, M. The nonlinear biological dynamics of unidirectional rate sensitivity. Annals of N.Y. Acad. of Sci., 98(4):806, 1962.
8. Coleridge, J.C.G. and R.J. Linden. The effect of infusion upon the heart rate of the anesthetized dog. J. Physiol., 128:310, 1955.
9. Defares, J.G., J.J. Osborn and Hiroshi H. Hara. Theoretical synthesis of the cardiovascular system. Study I: The controlled system. Acta Physiol. Pharmacol. Neerl., 12:189, 1963.
10. Ead, H.W., J.H. Green and E. Neil. A comparison of the effects of pulsatile and non-pulsatile blood flow through the carotid sinus on the reflexogenic activity of the sinus baroreceptors in the cat. J. Physiol., 118:509, 1952.
11. Guyton, A.C. and J.W. Harris. Pressoreceptor-autonomic oscillation: A probable cause for vasomotor waves. Am. J. Physiol., 165:158, 1951.
12. Grodins, F. Control Theory and Biological Systems. Columbia University Press, 1963.

13. Jewett, D.L. Activity of single efferent fibers in the cervical vagus nerve of the dog with special reference to possible cardio-inhibitory fibers. *J. Physiol.*, 175: 321, 1964.
14. Katona, P.G., G.O. Barnett, and W.D. Jackson. An analysis of heart rate control. *Fed. of Am. Soc. for Exper. Biol. Proc.* 22:182, 1962.
15. Katona, P.G. Computer Simulation of the Blood Pressure Control of Heart Period. Sc.D. Thesis, Dept. of Elec. Eng., M.I.T. 1965.
16. Katona, P.G. Computer simulation of the blood pressure control of heart period. *Proc. of the 18th Annual Conf. on Engineering in Med. and Biol.*, Philadelphia, 1965.
17. Kinnison, G.L. Heart Rate and Respiration Control Reflexes Induced by Systemic and Pulmonary Blood Pressure Changes. Ph.D. Thesis in Physiology, Univ. of Calif., Los Angeles, 1963.
18. Koepchen, H.P., H.D. Lux and P.H. Wagner. Untersuchungen über Zeitbedarf und Zentrale Verarbeitung des Pressoreceptorischen Herzreflexes. *Pflugers Arch.*, 273:413, 1961.
19. Landgren, S. On the excitation mechanism of the carotid baroreceptors. *Acta Physiol. Scand.*, 26:1, 1952.
20. Lercari, R.F. Mathematical Relation Between Baroreceptor Neural Activity and Arterial Pressure. S.M. Thesis, Dept. of Elec. Eng., M.I.T. 1964.
21. Lercari, R.F., G.O. Barnett and W.D. Jackson. Linear system analysis of blood pressure regulation. 17th Annual Conf. on Engineering in Med. and Biol., Cleveland. Nov. 16-18, 1964.
22. Levison, W.H. and G.O. Barnett. Nonlinear analysis of the pressoreceptor reflex system. *Proc. of the 16th Annual Conf. on Engineering in Med. and Biol.*, Baltimore. 1963.
23. Levison, W.H. Nonlinear Analysis of the Pressoreceptor Reflex System. Sc.D. Thesis, Dept. of Elec. Eng., M.I.T. 1964.
24. McCubbin, J.W., J.H. Green and F.H. Page. Baroreceptor function in chronic renal hypertension. *Circ. Res.*, 4:205, 1956.
25. McDonald, D.A. Blood Flow in Arteries. Edward Arnold Ltd., London. 1960.

26. Peiss, C.N. Concepts of cardiovascular regulation: past, present and future. Nervous Control of the Heart (ed. Randall), The William and Wilkins Co., Baltimore, p. 154, 1965.
27. Peterson, L.H. Systems behavior, feed-back loops, and high blood pressure research. Circ. Res., 12:585, 1963.
28. Pieper, H.P. Adaptation of cardiac output to peripheral runoff studied in intact dogs. Science 131:828, 1960.
29. Pruslin, D.H. Neural Patterns in Blood Pressure Regulation. S.M. Thesis, Dept. of Elec. Eng., M.I.T. 1964.
30. Rosenblueth, A. and F.A. Simeone. The interrelations of vagal and accelerator effects on the cardiac rate. Am. J. Physiol., 110:42, 1934.
31. Roston, S. Blood pressure and the cardiovascular system. Annals of the N.Y. Acad. of Sciences, 96:962, 1962.
32. Rushmer, R.F., O.A. Smith and D. Franklin. Mechanisms of cardiac control in exercise. Circ. Res., 7:602, 1959.
33. Sarnoff, S.J., J.P. Gilmore, S.K. Brockman, J.H. Mitchell and R.J. Linden. Regulation of ventricular contraction by the carotid sinus. Circ. Res., 8:1123, 1960.
34. Schaefer, H. Central control of cardiac function. Physiol. Rev., Suppl. 4, 40:213, 1960.
35. Scher, A.M. and A.C. Young. Servoanalysis of carotid sinus reflex effects on peripheral resistance. Circ. Res., 12: 152, 1963.
36. Stegemann, J. Der Einfluss sinusformiger Druckänderungen im isolierten Karotissinus auf Blutdruck und Pulsfrequenz beim Hund. Verhandlungen Deutschen Gesellschaft für Kreislaufforschung, 23:392, 1957.
37. Teorell, T. Some biophysical considerations on pressoreceptors. Arch. Int. Pharmacodyn., CXL 3-4, 563, 1962.
38. Trank, J.W. Studies on the Carotid Sinus Baroreceptors with Particular Reference to Endotoxin Shock. Ph.D. Thesis in Physiology, Univ. of Minnesota, 1961.
39. Turner, L.D. Peripheral Resistance Dependence upon Sympathetic Neural Stimulation in the Hindlimb of the Dog. S.M. Thesis, Dept. of Elec. Eng., M.I.T. 1965.

40. Van der Pol, B. Le battement du coeur considere comme oscillation de relaxation et un modele electrique du coeur. *L'onde Electrique* 7:365, 1928.
41. Warner, H.R. Use of analog computers for analysis of control mechanism in the circulation. *Proc. IRE* 41:1913, 1959.
42. Warner, H.R. and A.F. Toronto. Regulation of cardiac output through stroke volume. *Circ. Res.* 8:549, 1960.
43. Warner, H.R. and A. Cox. A mathematical model of heart rate control by sympathetic and vagus efferent information. *J. Appl. Physiol.* 17:349, 1962.
44. Warner, H.R., W.S. Topham and K.K. Nicholes. The role of peripheral resistance in controlling cardiac output during exercise. *Annals N.Y. Acad. of Sciences*, 115:669, 1964.
45. Warner, H.R. Simulation as a tool for biological research. *Simulation*, 1:57, 1964.
46. Weidinger, M., R. Hetzel and M. Schaefer. Aktionsstrome in zentrifugalen vagalen Herznerven und deren Bedeutung fur den Kreislauf. *Pflugers Arch.*, 276:262, 1962.
47. Topham, W.S. Control of Cardiac Output Studied with Computer Techniques. Ph.D. Thesis, Dept. of Biophysics and Bioengineering, Univ. of Utah.

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Technical information generated in connection with a NASA contract or grant and released under NASA auspices.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other nonaerospace applications. Publications include Tech Briefs; Technology Utilization Reports and Notes; and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546